PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, 15/86, C07K 14/47, C12Q 1/68

A1

(11) International Publication Number: WO 95/20044

(43) International Publication Date: 27 July 1995 (27.07.95)

(21) International Application Number: PCT/GB95/00095

(22) International Filing Date: 19 January 1995 (19.01.95)

(30) Priority Data:

9400929.7	19 January 1994 (19.01.94)	GB
9422021.7	31 October 1994 (31.10.94)	GB

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ).

Published
With international search report.

(71) Applicant *(for all designated States except US)*: LYNXVALE LIMITED [GB/GB]; The Old Schools, Trinity Lane, Cambridge CM2 1TS (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: BARTON, Charles, Howard [GB/GB]; 9 Hillcrest Gardens, Waltham Chase, Southampton SO32 2NF (GB). WHITE, Jacqueline, Katie [GB/US]; 1 Exeter Park #1, Cambridge, MA 02140 (US) BLACKWELL, Jenefer, Mary [GB/GB]; 19 Quakers Lane, Potters Bar, London EN6 1RL (GB).

(74) Agents: DANIELS, Jeffrey, Nicholas et al.; Page White & Farrer, 54 Doughty Street, London WC1N 2LS (GB).

(54) Title: NATURAL RESISTANCE ASSOCIATED MACROPHAGE PROTEIN AND USES THEREOF

(57) Abstract

Macrophages, the main phagocytic cells of animals, can bind and ingest microorganisms. A natural resistance-associated macrophage protein is provided having an N-terminal region comprising an Src homology 3 binding domain. When present in the macrophage, the protein is capable of controlling resistance to the microorganisms.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
				TJ	Tajikistan
FR	France	**MN**	Mongolia		
GA	Gabon				

NATURAL RESISTANCE ASSOCIATED MACROPHAGE PROTEIN AND USES THEREOF

FIELD OF THE INVENTION

The present invention relates to a nucleotide sequence encoding a natural resistance-associated macrophage protein, the protein product thereof, nucleotide probes and primers thereto, polypeptide fragments of the protein and related antibodies.

BACKGROUND TO THE INVENTION

Macrophages are the main phagocytic cells of animals and play a key role in the immune system. Macrophages bind and injest particles recognised as foreign by the immune system. Such particles include microorganisms.

The three microorganisms *Salmonella typhimurium*, *Leishmania donovani* and *Mycobacterium bovis* (BCG) are all intracellular pathogens of macrophages. Three separate groups of scientists had previously identified genes capable of controlling resistance and susceptibility to each of these microorganisms. The genes were designated respectively Ity, Lsh and Bcg. Subsequent work has led the scientists to conclude that Ity/Lsh/Bcg is a single gene and is expressed at the macrophage level (Ref 1).

Recently, Vidal *et al* (Ref 2) cloned a murine gene as the most likely candidate to be Lsh/Ity/Bcg. This gene has been termed the natural resistance-associated macrophage protein (Nramp) gene. A cDNA for Nramp was isolated from a pre B-cell cDNA library and sequenced. The amino acid sequence for the protein product was deduced from the nucleotide sequence and predicts a 53kDa protein. On the basis of the deduced amino acid sequence, Vidal *et al* proposed as a function of the Nramp protein the transport of nitrate across the membrane of the intracellular

via nitrite to toxic nitric oxide thereby to enhance killing of

the microorganisms.

The present applicants have isolated and sequenced a macrophage-expressed Nramp cDNA. Contrary to the teaching of Vidal *et al* the present applicants have found a different nucleotide sequence including a region encoding an additional amino acid sequence at the N-terminus. Surprisingly, the additional amino acid sequence includes structural features which may be responsible for protein-protein interactions essential in signal transduction pathways thereby suggesting that Nramp controls early amplification of transmembrane signalling in disease resistant macrophages by binding the SH3 domain of tyrosine kinases or other molecules.

SUMMARY OF THE INVENTION

The present invention provides in one aspect a natural resistance-associated macrophage protein having an N-terminal region comprising an SH3 binding domain. When present in the macrophage, the protein is capable of controlling resistance to pathogenic microorganisms.

SH3 (Src homology 3) domains are believed to mediate specific protein-protein interactions required in signal transduction (Ref 3) and have been identified as related sequences in a variety of proteins (Refs 4 and 5). In one embodiment of the present invention, the SH3 binding domain comprises the SH3 binding motif PGPAPQPXPXR, more particularly PGPAPQPAPCR. This motif is found in the protein obtainable from mice. In another embodiment of the present invention, the SH3 binding domain comprises the SH3 binding motif PXSPTSPXPXXAPPRXT, more particularly PTSPTSPGPQQAPPRET. This motif is found in the protein obtainable from humans. Typically, SH3 binding domains are rich in proline and sometimes serine.

at the N-terminal end of the SH3 binding

particularly, the polypeptide segment is SPPRLSRPSYGSISSL. The SH3 binding domain obtainable from humans preferably further comprises the polypeptide segment GPQRLSGSSYGSISS.

A further preferred feature of the N-terminal region is the presence of one or more consensus sequences for protein kinase C (PKC) phosphorylation. Preferably, the N-terminal region has two protein kinase C sites which flank the SH3 binding domain. Tyrosine residues may also flank the SH3 binding domain.

Typically, the N-terminal region comprises 64 amino acids.

The full amino acid sequences of the murine and human proteins are set out in Figure 9. Mutations or deletions may be present in each sequence provided that they do not substantially affect the activity of the protein.

In a second aspect, the present invention provides a nucleotide sequence encoding the natural resistance-associated macrophage protein discussed above. Where the nucleotide sequence is a DNA sequence, this may be a genomic sequence containing introns and exons or a cDNA sequence obtainable from mRNA by reverse transcription. The SH3 binding domain of the protein obtainable from mice is preferably encoded by the DNA sequence comprising CCTGGCCCAGCACCTCAGCCAGCGCCTTGCCGG and may further comprise the upstream region AGCCCCCCGAGGCTGAGCAGGCCCAGTTATGGCTCCATTTCCAGCCTG. More particularly, the 5' end of the genomic DNA sequence is set out in Figure 4 and discussed in further detail below. A cDNA sequence is also provided, as set out in Figure 2. The SH3 binding domain of the protein obtainable from humans is preferably encoded by the DNA sequence comprising CCG ACC AGC CCG ACC AGC CCA GGG CCA CAG CAA GCA CCT CCC AGA GAG ACC
and may further comprise the upstream region GGT CCC CAA AGG CTA AGC GGG TCC AGC TAT GGT TCC ATC TCC AGC.

Various modifications and deletions [illegible]

sequences described may still result in a functional protein product. Owing to the degeneracy of the genetic code it will be readily apparent that numerous silent mutations within the specified sequences will give rise to the same amino acid sequence.

The cDNA sequence has been deposited as part of plasmid pBabeλ8.1 under accession number NTC 12855 at the National Collection of Type Cultures, Central Public Health Laboratory, London, UK. The deposit consisted of a culture of *E. Coli* DH5α transformed with the plasmid and was given the date of deposition of 14 January 1994.

In a further aspect, the present invention provides a retroviral vector construct incorporating a cDNA sequence encoding the natural resistance-associated macrophage protein.

Generally, the use of retroviral vectors presents a good method for gene transfer into haematopoietic cells and has advantages over other methods including stable transfer of a single copy of a gene into the recipient cell, and high efficiency of gene transfer into target cells. Vector constructs are generated by ligating the gene of interest using standard molecular biology techniques into a non-replication-competent viral genome. The resultant vector construct is transferred into a cell line (the packaging cell line) capable of replicating the viral genome and packaging it into infective pseudovirus particles. Depending on the virus envelope protein encoded by the packaging cell, the resulting pseudovirus particles can be capable of infecting a wide host range (amphotropic) or be restricted to rodent cells (ecotropic).

In the present application the pBabe plasmid was used as a suitable retroviral vector. This plasmid is discussed in further

marker gene which confers antibiotic resistance. Antibiotic resistant clones can then be tested for their ability to secrete functional pseudovirus particles and to infect recipient cells.

The selected retroviral constructs of the gene carrying Nramp can be used as the basis for gene therapy to create a functional copy of the gene where lack of expression has been observed, or where a non-functional copy exists. The most likely method of gene transfer is via the bone marrow or progenitor cells isolated from the circulation. The gene can be retrovirally introduced into these stem cells removed from the patient. The stem cells would then be reintroduced into the patient with the aim of repopulating the myeloid/lymphoid cell lineages with cells containing a functional copy of the gene.

In a further aspect, the present invention provides nucleotide probes or primers capable of hybridizing to a portion of the nucleotide sequence described, preferably to at least a portion of the sequence above which encodes the N-terminal region of the protein comprising or upstream of the SH3 binding domain. Probes can be single or double stranded and can be made by recombinant DNA technology from copies of the gene, or portions thereof, or by synthetic routes such as which lead to oligonucleotide probes. Primers, such as those for polymerase chain reaction work, are single stranded and preferably are at least 18 nucleotides long. Both probes and primers based on the sequence can be used at both DNA and RNA levels for diagnosis and such probes and primers can be readily made using the sequence information provided herein. The cDNAs as described above are themselves useful probes for the gene. The present applicants have found that the cDNA sequence of the murine gene described above can be successfully used as a probe for the corresponding human gene.

The probes or primers have diagnostic potential, for example to

expressed but sub-functional or non-functional protein. Genetic defects to be diagnosed may occur in the coding sequence, or in the promoter or 3' untranslated regulatory regions of the gene and so probes directed to both genomic and cDNA sequences may be useful.

Primer pairs are also provided which are capable of hybridising to specific sequences in the 5' region of the human NRAMP gene, permitting amplification of a portion of the promoter region of the human gene. The promoter region preferably includes a poly gt site, especially in the configuration $t(gt)_5ac(gt)_5ac(gt)_ng$ in which n=o, or an integer. The promoter region may further comprise a transcription site, and, optionally, one or more of: an Interferon-Y response element; a NFKB site; an AP-1 site; a W-element; a PV.1 core motif; and a PEA3 site. Preferably each of the sites, elements or motifs are present in the order specified in Figure 11. Polymorphisms in the poly gt site, specifically located in the third cluster of gt repeats, where n may equal any number of repeats, typically 4 to 12, may be diagnostic for reduced or defective expression of the human gene and so primers permitting PCR amplification of this region are of particular importance. Probes to the promoter region are also provided, preferably allele-specific probes to the promoter region, for example allele-specific oligonucleotides.

In a preferred embodiment, the human protein is encoded by 15 exons, each of which is flanked by intron boundary regions. The exons are preferably those shown in Table 3. Probes or primers are provided, which are capable of hybridising to at least a portion of an individual exon and/or its flanking intron boundary region. Preferably, primer pairs are provided which are capable of hybridising to the intron boundaries of each exon so as to amplify the respective exon. More preferably, the primer pairs are capable of hybridisation to any one of the intron boundary

diagnostic of a defective gene

PCR amplification of exons to identify such polymorphisms using electrophoretic techniques are of particular importance. More preferably, the primer pair capable of hybridisation to the intron boundaries 5' and 3' of human exon 2, permitting PCR amplification of this exon, may be critically important in permitting detection of a polymorphism involving a 3 amino acid deletion in the putative SH3 binding domain encoded within this exon. The importance of this region to the function of the NRAMP gene represents a key component of this invention.

Antisense oligonucleotides may also be produced using the nucleotide sequence described above. Antisense oligonucleotides may be used to interrupt the expression of the gene and this could provide a potentially important local therapy for autoimmune disorders or cancers.

As discussed in further detail below, the protein product from the gene is predicted to be a polytopic membrane protein. Whilst antibodies against the protein will be important tools in diagnosing levels of expression of the protein product in various cell populations, only those portions of the protein which are not occluded by membrane are likely to be accessible to antibodies in the intact or native protein conformation. Accordingly, in a further aspect of the present invention there is provided a polypeptide fragment of the protein which comprises at least a portion of the structural domain not hidden by membrane. Preferably, the polypeptide fragment comprises at least a portion of the N-terminal region. Two structural domains of potential importance are: the N-terminal cytoplasmic domain proximal to the first membrane-spanning domain and comprising amino acids 1 to 82; and the C-terminal cytoplasmic domain distal to the last membrane-spanning domain and comprising amino acids 414 to 458.

Nramp are ligated. The pGEX series of prokaryotic expression vectors is a particularly useful type of vector into which the Nramp sequences may be ligated. This is a standard procedure, further information about which may be found in (Ref 8).

In a further aspect, the present invention provides an antibody to the natural resistance-associated macrophage protein or an antibody to a polypeptide fragment therefrom, more particularly to one of the accessible polypeptide domains discussed above. The Nramp fusion proteins may be used as antigens to innoculate rabbits or rats so as to produce antibodies. Using standard techniques both polyclonal and monoclonal antibodies may thus be raised.

In particular, antibodies recognising epitopes within specific amino acid sequences contained within the N-terminal (DKSPPRLSRPSYGSISS; PQPAPCRETYLSEKIPIP; and GTFSLRKLWAFTGPGFLMSIAFLDPGNIESDLQ) and C-terminal (WTCCIAHGATFLTHSSHKHFLYGL) regions of the protein will recognise the protein in both mouse and man, and can be applied for both research purposes and as a diagnostic tool in man.

In addition to diagnosis as discussed above, neutralizing monoclonal antibodies could be produced to block the function of the gene in situations where adverse effects are observed, such as autoimmunity or cancer resulting from expression of the gene.

The presence or absence of the gene product could have both beneficial and detrimental effects depending on the disease status. In infectious diseases, particularly involving intracellular pathogens of the myeloid cell lineage, absence of a functional gene product may result in chronic susceptibility to the disease. In the case of autoimmune disorders or cancers of the myeloid or lymphoid cell lineages, overexpression of the

diagnostic procedures and therapeutic agents associated herein may

be useful for patients presenting with atypical responses to infection, certain autoimmune disorders, or cancers of the myeloid or lymphoid lineages.

Another situation in which a deficit in the NRAMP gene might relate to cancer is where cancers of other cell lineages are destroyed by activated macrophages, through sensitivity to hydrogen peroxide generated by a respiratory burst response, TNF-α, or nitric oxide. All of these macrophage functions are regulated by NRAMP. In this case, corrective gene therapy via stem cell gene transfer would be appropriate.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be described further by way of example only with reference to the attached drawings in which:

FIGURE 1 shows a restriction map of the Nramp λ8.1 cDNA clone of the present invention as compared with that of Vidal *et al*;

FIGURE 2 shows the sequence of macrophage-λ8.1 Nramp in accordance with the present invention;

FIGURE 3 is a schematic representation of the genomic DNA corresponding to nucleotides 31 to 456 of λ8.1 as compared with the corresponding DNA of Vidal *et al*;

FIGURE 4 shows the 5′ sequence of the genomic DNA up to nucleotide 1911;

FIGURE 5(a) shows the results of northern blot hybridizations;
FIGURE 5(b) shows the results of primer extension on total RNA from B10.L-Lsh macrophages;

FIGURE 7 shows the nucleotide and deduced amino acid sequence of exon 2 of human NRAMP;

FIGURE 8 shows the result of an amino acid database search with the N-terminal sequence for human NRAMP;

FIGURE 9 shows the result of a Clustal V multiple sequence alignment for the deduced amino acid sequence for human NRAMP, murine *Nramp* clone λ8.1 [55], and the yeast mitochondrial proteins SMF1 and SMF2 [35];

FIGURE 10(A) shows the results of amino acid database searches for human NRAMP exon 2;

FIGURE 10(B) shows the results of a Clustal V multiple sequence alignment for human NRAMP, mouse *Nramp*, SMF1 and SMF2, and the expressed sequence tags [60] of *Oryza sativa* (rice; accession number d15268) and *Arabidopsis thaliana* (accession number z30530) genes, reading frames 1 and 2 respectively;

FIGURE 11 shows the 440 bp of putative promoter region human NRAMP sequence 5' of the transcription start site; and

FIGURE 12 shows two families segregating for (a) alleles 2 and 3, or (b) alleles 1, 2 and 3 of the 5' dinucleotide repeat polymorphism and autoradiographs of polymorphic PCR products separated by denaturing polyacrylamide gel electrophoresis.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Details of Experimentation and Results

Sequence-analysis of Nramp clones from macrophage cDNA library. Macrophage Nramp clones were isolated from an activated (4 h stimulation; 25 U/ml interferon-γ, 10ng/ml Salmonella typhimurium

corresponding to nucleotides 1410 - 1812bp of the published (Ref 2) sequence. Following plaque purification 35 clones from 10^6 recombinants were analysed by PCR using sense and anti-sense Nramp primers in combination with T3 and T7 vector arm primers. This allowed the mapping of clones with respect to the published sequence. 20/35 were found to be 1.0 - 1.5 kb and were not analysed further. The remainder were of 2.1 - 2.3 kb and potentially encoded full length Nramp coding sequence. Clones were initially restriction mapped and four selected for sequencing (Sequenase II) including the longest clone λ8.1.

Genomic sequencing.

From the macrophage cDNA sequence, PCR primers were generated to amplify a 2kb region of DNA from both yeast artificial chromosome (YAC clone C9C28; Princeton library) and mouse genomic DNAs. The products were cloned in the pCR vector (Invitrogen) and sequence (Sequenase II) determined from double stranded plasmid DNA from at least two clones of each, using oligos complementary to the cDNA sequence. Splice junctions were identified by comparison of the genomic and complementary DNA sequences.

Northern blot and primer extension analysis of Nramp expression.

Cytoplasmic total RNA isolated in the presence of vanadyl ribonucleoside complexes was utilised for denaturing gel electrophoresis with glyoxal and Northern blotting, or directly for RT reactions for primer extension analysis. Hybridizations were performed using probes isolated from a genomic fragment (bp 1 - 1482) 5' of exon 3 (see results). Restriction digestion with BamHI generated two probes covering a λ8.1-specific (= bp 1 - 587 of the genomic sequence; Fig. 2b) region or the putative 5' untranslated sequence (= bp 588 - 1482 of the genomic sequence; Fig. 2b) of the published (Ref 2) cDNA. For primer extensions oligonucleotides designed to be specific for λ8.1-like RNAs (TCT GCG CTG GGA ATG GGG; bp 538-521 of the genomic sequene) or for

with 25 Units of AMV reverse transcriptase at 42°C, terminated by the addition of gel loading buffer, and sized against a sequencing ladder following denaturing polyacrylamide gel electrophoresis.

Referring to Figure 1, cDNA clones isolated from an activated macrophage library carrying the resistant allele of murine Nramp were restriction mapped, sequenced and found to be identical over the coding region of the published (Ref 2) sequence, except for two silent mutations (359bp, C; 965bp, T). Regions of sequence identity between the published clone and macrophage clone λ8.1, which includes the ATG (d) codon and the major ORF (solid bars) of the published clone, are shown within the broken lines. The positions of SmaI (S) and PvuII (P) cleavage sites demonstrate divergence between the clones at the 5' end. Novel sequence identified in the 5' region of λ8.1 contained a more proximal ATG (p) codon and an extended ORF (open bar) encoding an additional 64 N-terminal residues compared with the published Nramp.

Figure 2 shows the sequence of macrophage λ8.1. The nucleotide sequence specific to λ8.1 is underlined. The 64 N-terminal residues encoded by λ8.1 occur 5' of the distal initiation codon (=Met at position 65) identified by Vidal and coworkers (Ref 2). This additional 64 amino acid sequence is identical in resistant and susceptible mice (data not shown) and is rich in Ser 10/64, Pro 10/64, basic 7/64 residues, and contains 3 consensus PKC phosphorylation sites (S/T-X-R/K) on Ser 3, 37, and 52. As described previously (Ref 2), putative N-linked glycosylation sites occur at residues 311 and 325, and hydrophobic potential membrane-spanning domains are underlined. Database searches also revealed a B-2 alu-like repetitive element (boxed) within the 3' UTR, which produces complex signals when the full-length λ8.1 clones is hybridised to mouse genomic DNA.

nucleotides 31-456 of λ8.1 and spanning the point of divergence with the published (Ref 2) clone was isolated and sequence determined to elucidate the mechanism generating the two clones. The additional sequence of macrophage Nramp is encoded by two unique exons (1 and 2; solid bars) contiguous in the cDNA sequence (figure 1) with exons (3 and 4; solid bars) common to both λ8.1 and the published clone. Contiguous with and 5' of the third exon is the putative 5' UTR (open bar) found in the published clone. Predicted splicing patterns (dotted line) are indicated above (λ8.1) and below (published pre-B cell clone) the map. Also shown are the sequencing gel reads (arrows) determined using specific Nramp primers from the cDNA sequence, as well as new primers specific to the genomic sequence.

Figure 4 shows the sequence of genomic DNA spanning the point of divergence of λ8.1 and the published (Ref 2) sequence. Exonic nucleotide sequence is shown in capitals, with the predicted amino acid sequence indicated above in single letter format. Intron sequence is shown in small letters. The region of 5' UTR from the published clone, contiguous with the third exon, is highlighted by overlining. The codon (ATG = Met) where this terminates indicates the initiation codon of the published sequence. A probe containing sequence unique to the 5' region (bp 1 - 587) of the mouse genomic sequence also hybridises to genomic Lamda clones isolated from a subcloned human YAC (clone AM11/D3/14; ICRF library) known to hybridise to the homologous region of human 2q35.

Figure 5 shows that Nramp transcripts encoding the additional 64 amino acids are the only form of Nramp expressed in the macrophage.

Referring to Figure 5(a), to identify the nature of Nramp

lanes 1,2,3,4, of macrophages activated with ribolipid

(lanes 2,6,10,14), LPS (lanes 3,7,11,15), or interferon-γ plus LPS (lanes 4,8,12,16). Probes specific for unique λ8.1 5'sequence (lanes 1-4), or for the more distal putative 5'UTR of the published (Ref 2) sequence (lanes 5-8), were used and compared against the same blots reprobed with constitutively expressed GAPDH (lanes 9-16). Hybridizing RNAs could only be detected with the λ8.1-specific sequence, despite loading twice as much RNA on blots hybridised with the more distal probe. Results are shown for RNA extracted from bone marrow-derived macrophages from C57BL/10ScSn mice. Slot blot analysis (not shown) confirmed that the λ8.1-specific probe hybridized to RNA from both susceptible and resistant macrophages. Southern blot analysis (not shown) confirmed that both probes hybridized to EcoRI fragments of 3500 and 500 bp in mouse genomic DNA from both C57BL/10ScSn (Lsh^s) and B10.L-Lsh^r mice.

Referring to Figure 5(b), to identify the 5'terminus of Nramp transcript expressed in macrophage RNA, primer extensions were performed with 10μg of total RNA from B10.L-Lsh^r macrophages using oligonucleotides specific to the putative 5' untranslated region of the published (Ref 2) sequence (lane 1), or to 5' sequence unique to λ8.1 (lane 2). The numbers of nucleotides from the 5' end of the primer are shown. Control reactions with tRNA gave no products with either primer. These experiments confirm that RNA transcripts bearing the putative 5' untranslated region of the published cDNA are not present in resting (not shown) or activated macrophages, whereas transcripts corresponding to the λ8.1 sequence were identified with transcriptional initiation sites mapping 21 and 22bp (doublet) 5' of the proximal ATG codon. Similar results were obtained using RNA from C57BL/10ScSn (*Lsh*^{s}) macrophages as template.

Figure 6 shows that macrophage Nramp encodes an N-terminal SH3 binding domain structure.

sequence unique to macrophage-expressed Nramp

of sequence matches particularly with the proline rich sequence. Multiple sequence alignments allowed for the generation of a consensus motif over this region: PGPAPQPXPXR (solid vertical bar). Matches were found for three molecules involved in signal transduction: the focal adhesion kinase (Ref 20) (50% identity over 26 residues); Drosophila dynamin shibire protein (Ref 17) (55% identity over 20 residues); and the adenylate cyclase stimulatory beta-1-adrenergic receptor (Ref 19) (57% identity over 21 residues). A proline, serine rich domain has been identified as a functional SH3 binding domain in dynamin (Refs 17,3). The nine best matches were aligned with each other and residues boxed where four or more exhibited identities. Also shown are the two PKC sites (hatched vertical bars) on S3 and S37 which flank the region exhibiting sequence identity. Tyrosine residues (asterisk) occur on either side of the consensus motif indicating conservation of this part of the sequence.

Figure 7 shows the nucleotide and deduced amino acid sequence of exon 2 of human NRAMP obtained from genomic sequence analysis.

Exon 2 in murine Nramp encodes the putative SH3 binding domain with amino acid matches to a number of signal transduction molecules. To characterise the structure of the same region of the human NRAMP gene, a yeast artificial chromosome hybridising to 2q35 was subcloned into EMBL3 and the resulting library screened by cross-species hybridisation using a murine \probe to identify clones containing this exon. Genomic sequence across exon 2 was obtained (figure 7), with splice donor and acceptor sites conforming to the GT AG boundaries as identified in the murine sequence.

Figure 8 shows that N-terminal sequence for human NRAMP encodes an SH3 binding domain structure.

sequence maintained (Figure 8). These include the distal consensus sequence for phosphorylation by PKC. Both tyrosine (Y) residues are maintained and share identical positions as murine Nramp, and the human exon 2 sequence is rich in serine (S 9/48 compared to 9/45 in mouse); proline (P 10/48 compared to 10/45 in mouse); and basic (5/48 compared to 5/45 in mouse) residues. Of these important residues, 6/9 S, 6/10 P, and 4/5 basic residues show identical positions within murine and human exon 2. The spacing of the prolines are subtly different in the consensus sequence for the SH3 binding domain of the human gene: at positions 1 4 7 9 13 14 compared with 1 3 5 7 9 in mouse. The human consensus motif over this region is: PXSPTSPXPXXAPPRXT. The 3 codon insertion in human exon 2 forms the 5' segment of this proline rich domain. This insertion region has an unusual nucleotide sequence consisting of an almost perfect 3 times 9 nucleotide repeat, representing a region of some instability and source of polymorphism in man (Ref 32) which could influence function. The presence of the extra 3 codon segment within the human gene sequence produced some additional amino acid sequence identities on screening databases. These include several proteins involved in cytoskeletal interactions or signal transduction pathways: microtubule associated protein 4; adenylyl associated protein; phospholipase C β3; phosphatidylinositol 3-kinase regulatory subunit p85α (PI3-kinase p85α); ankyrin; and zyxin.

Computer-assisted analysis.

Hydropathy profiles of the predicted N-terminal amino acid sequence of macrophage-expressed Nramp were obtained by computer-assisted analysis using the algorithm and hydropathy values of Kyte and Doolittle (Ref 14). Amino-acid sequence comparisons were made using the FASTA programme on-line to the CRC Resource Centre.

Nramp cDNAs, 14 of which differed from the published (Ref 2) Nramp in the 5' terminal sequence. The longest macrophage-derived cDNA (Fig. 1; λ8.1) was 186bp shorter than Nramp. It contained the full length coding region for the previously predicted protein, and exhibited 100% identity with no in frame stop codons for the region (bp 209-263) of untranslated sequence immediately 5' of the published initiation codon. However, nucleotides 1-208 of λ8.1 shared no identity with the published sequence. A more proximal ATG codon was identified at 72bp in λ8.1, preceeded by an in frame stop codon at 36bp. This proximal translational initiation codon is followed by an ORF of 192bp (64 amino acids) that leads into the ORF previously reported. Previous studies have shown that proximal initiation codons are utilised in more than 90% of all genes analysed (Ref 15). Nor was there any evidence that the distal initiation codon would be favoured, since both distal and proximal initiation codons and flanking sequences are identical (TCCTCATGA) and display only two identities with the optimal (Ref 16) ($CC^{A}/_{G}CCATGG$) consensus. Hence, there is no a priori reason why the distal initiation codon would be used.

Genomic sequence for the 5' region of Nramp. To determine whether mechanisms exist which could generate two RNAs and hence two types of Nramp clones, a region of genomic DNA spanning the point of divergence was characterised corresponding to nucleotides 31-456 of λ8.1 (Fig. 2). This region is encoded by four exons interspersed by three introns of 395, 900 and 241bp, with all splice donor and acceptor sites conforming to the GT and AG boundaries. The first 47 amino acids of the 64 amino acid N-terminal domain of λ8.1 are encoded by two proximal exons unique to this clone. The remaining 17 amino acids are encoded by exon three, with exons three and four common to both λ8.1 and the published (Ref 2) Nramp cDNA. The 5' UTR sequence from the published clone was found in the 900bp intron contiguous with and

published Nramp sequence it contains both coding and non-coding sequence. Although a complex mechanism involving alternative splicing associated with an internal splice acceptor site and dual promoter control could be formulated to describe the origin of both forms, it seems more likely that the published (Ref 2) cDNA clone contains a fragment of the 900bp intron at its proximal end. This is consistent with the observation that a number of the macrophage-derived Nramp clones isolated here were found to contain sequence that exhibited identity with the first Nramp intron identified in genomic DNA (not shown).

Only one form of Nramp is expressed in macrophages. To confirm the hypothesis that the RNA encoding the longer polypeptide is the form expressed in macrophages, a number of different experimental approaches were adopted (Fig. 3). Using macrophage RNA as template, primer extension with an oligonucleotide unique to the 5′ region of λ8.1 yielded products in both susceptible and resistant mice. In direct contrast, no products were generated using an oligonucleotide within the putative 5′ UTR of published (Ref 2) Nramp. A probe covering the 5′ region unique to λ8.1 also hybridized well to Northern and slot blots of macrophage RNA from susceptible and resistant mice, whereas a probe covering the putative 5′ UTR of the published clone showed no hybridization. Hence, the only form of RNA transcript present in macrophages is that which conforms to the λ8.1 predicted polypeptide sequence, suggesting that this form of the Nramp gene is responsible for host resistance.

Predicted structure and sequence identities across the N-terminus of macrophage-expressed Nramp. In order to determine how macrophage-expressed Nramp might relate to Lsh/Ity/Bcg gene function, hydropathy (Ref 14) plots and amino acid database searches were undertaken over the newly identified 64 amino acid domain. The former (data not shown) demonstrated that the new

three PKC phosphorylation sites (in addition to the two identified in the published Nramp sequence), and a number of matches with several unrelated proteins (Fig. 4). The most intriguing matches were: (i) with the dynamin shibire protein (Ref 17) of Drosophila, related to mammalian dynamin (dephosphin) which acts as a synaptic phosphoprotein in rat brain (Ref 18); (ii) with the proline rich third cytoplasmic domain of the adenylate cyclase stimulatory and G protein coupled beta-1-adrenergic receptor (Ref 19); and (iii) with the focal adhesion kinase (Ref 20) that can be modulated by integrin-dependent phosphorylation (Ref 21). The region of identity with the C-terminal domain of dynamin has been implicated (Ref 22) in binding anionic phospholipids, microtubules and Src homology 3 (SH3) domains. SH3 domains (Refs 3,4), identified as related sequences in different tyrosine kinases (TK) but outside the catalytic domain, are modular and found in a number of proteins such as the non-receptor TKs, phospholipase C-gamma and other structural proteins of the cytoskeleton. Whilst the function of SH3 domains (Ref 4) is not as well characterised as the SH2 counterpart, it is believed they mediate specific protein-protein interactions obligatory for signal transduction (Ref 3). IL-2R beta (Ref 25) and erythropoietin (Ref 26) receptors, for example, exhibit serine and proline rich intracellular domains which associate with TKs mediating phosphorylation essential to receptor function. Members of the Src family of membrane-associated TKs, including Hck and Fgr, are also found in macrophages (Ref 27). Both exhibit differential kinetics in response to priming/activation signals and could be implicated in Nramp-mediated signal transduction pathways. Hck, in particular, has recently been shown to be involved in signal transduction for TNF-α production in murine macrophages (Ref 28), a step which we have demonstrated (Ref 11) is crucial in the pathway to enhanced nitric oxide production and antimicrobial activity in Lsh resistant macrophages.

which mediate binding to extracellular matrix proteins and signal via TKs, is sufficient to stimulate enhanced TNF-α production in resistant but not susceptible macrophages (Ref 9). Overall, the multiple PKC phosphorylation sites on Nramp, together with the new SH3 binding domain identified here, provide compelling evidence that Nramp mediates resistance by controlling signal transduction for macrophage priming/activation.

N-terminal sequence analysis of human NRAMP supported the findings with murine Nramp in showing sequence identity over the putative SH3 binding domain with a series of proteins involved in cytoskeletal interactions or signal transduction pathways. Of these, PI3-kinase p85α (Ref 10) is of particular interest because it functions by binding to phosphorylated protein tyrosine kinase via SH2 domains (Ref 12), and acts as an adaptor mediating the association of the p110 catalytic unit to the plasma membrane. PI3-kinase p85α also has an SH3 domain. Ankyrin B (Ref 13) is a molecule linking integral membrane proteins to cytoskeletal elements, and zyxin (Ref 23), an adhesion plaque protein and a possible component of a signal transduction pathway mediating adhesion-associated gene expression. Overall, this evidence supports our earlier conclusion based on the putative SH3 binding domain of the murine gene that this domain is important in protein-protein interactions important in signal transduction, and/or protein interactions (e.g. binding of tyrosine kinases mediating phosphorylation on tyrosines) which regulate the transport function of the molecule.

Nramp gene transfer studies.

A number of Nramp retroviral vector constructs were made, all based on the pBabe plasmid. These include the cDNAs encoding the predicted protein described above, together with a C-terminal deletion construct encoding the proximal 72 amino acids of the

marker gene which confers resistance to puromycin. A number of these resistant clones have been tested for their ability: (i) to secrete functional pseudovirus particles by RNA slot blotting and hybridisation with an Nramp probe; and (ii) to infect recipient cells and confer antibiotic resistance. Infectious particles from the highest titre lines will be used for in vivo gene transfer. This same construct has been introduced into a murine macrophage cell line (RAW 264) which expresses a different allelic ("Lsh susceptible") variant from that of the vector-derived Nramp gene.

Several clones have been identified that co-express both forms of Nramp as monitored by PCR followed by allele-specific oligonucleotide hybridisation. Functional experiments have been performed to demonstrate that Nramp is the disease resistance gene Ity/Lsh/Bcg, by demonstrating that the resistant allele confers macrophage activation phenotypes previously associated with the action of the Ity/Lsh/Bcg gene. More specifically:-

Table 1 demonstrates that the Nramp resistant allele confers an enhanced baseline PMA-elicited respiratory burst response compared to the control susceptible transfectant clones. This resting PMA-elicited respiratory burst is completely extinguished in susceptible but not resistant transfectants following treatment of the macrophages with bacterial lipopolysaccharide (LPS). Respiratory burst products mediate antimicrobial and tumouricidal activity.

Table 1. Resistant allele RAW264.7 transfectants generate enhanced RB responses which are not extinguished following LPS stimulation. PMA-elicited RB was measured using a standard assay in which superoxide reduces nitro blue tetrazolium to formazan in (a) resting resistant and susceptible transfectants, and (b) after 24 or 30 hours incubation with LPS (25ng/ml). To normalise

$p<0.001$ = ***) for results of student's t-tests used to compare each resistant transfectant against the susceptible transfectant 30S. Similar levels of significance were observed for comparisons with 10S and 25S. Results representative of 5 independent experiments performed.

Transfectant	24 hour resting cells	24 hour LPS/IFNγ treated
10S	0.155±0.013	0.025±0.006
25S	0.181±0.026	0.002±0.003
30S	0.147±0.025	0.034±0.008
7.1R	0.296±0.056*	0.219±0.022***
7.2R	0.399±0.077***	0.292±0.029***
7.5R	0.442±0.080**	0.181±0.097*
7.8R	0.291±0.019**	0.364±0.052***
7.11R	0.290±0.069**	0.308±0.036***
17.1R	0.389±0.082*	0.272±0.059**
17.5R	0.329±0.056**	0.230±0.018***

Table 2 demonstrates that the Nramp resistant allele confers enhanced nitrite release following priming/activation with LPS and interferon-γ (IFNγ). Nitrites are the stable end-product of nitric oxide generated by upregulated expression of the inducible nitric oxide synthase gene in resistant macrophages. Nitric oxide also mediates antimicrobial and tumouricidal activity, and is specifically known to be the final effector mechanism for antileishmanial and antimycobacterial activity in murine resistant macrophages.

endproduct of NO production using the Griess reagent. Cells were incubated for 24 or 30 hours in the presence of LPS (25ng/ml) or LPS plus IFNγ (25 ng/ml and 25 U/ml). Determinations were normalised to cell number from the crystal violet staining intensity of a parallel plate and results are presented as the ratio of nitrite to crystal violet. Asterisks indicate significance levels ($p<0.05$ = *; $p<0.01$ = **; $p<0.001$ = ***) for results of student's t-tests used to compare each resistant transfectant against the susceptible transfectants 10S and 30S. Clones 17.3R and 17.6R developed from an independent transfection also showed significantly ($p<0.05$) higher NO levels in this experiment. Results representative of 5 independent experiments performed.

Transfectant	LPS alone	LPS + IFNγ
10S	0.043±0.008	0.296±0.026
30S	0.009±0.008	0.274±0.027
7.2R	0.215±0.005***	0.603±0.059**
7.5R	0.256±0.062***	0.857±0.059***

Table 3 demonstrates that the Nramp resistant allele confers enhanced L-arginine uptake following priming/activation with LPS and IFN-γ. L-arginine provides the substrate for generation of nitric oxide involved in signal transduction for upregulated expression of KC in resistant macrophages (Ref 24), and for the final effector mechanism for cidal activity of the macrophage.

Table 3. L-arginine uptake is enhanced in resistant transfectants compared to susceptible following activation with

contribution from serum. Pilot experiments demonstrated that the uptake of [^{3}H] L-arginine (0.25 μCi, spe cific activity 58 Ci/mmol) from 10^5 cells was linear over a one hour time period at 37°C. In all subsequent experiments cells were pulsed for 30 - 45 mins. The incubation was terminated by removing the media and washing the adherent cells 3 times in PBS containing 10 mM unlabelled L-arginine. Cells were lysed in 50 μl of 1% SDS and counted in 5 ml of aquasol II (DUPONT-NEN). Results are expressed as the percentage stimulation $\pm$ standard deviation observed in 6 hour LPS + IFNγ treated macrophages compared to untreated controls. Asterisks indicate significance levels ($p<0.05$ = *; $p<0.01$ = **; $p<0.001$ = ***) for results of student's t-tests used to compare each resistant transfectant against the susceptible transfectants 2S, 10S and 30S. Results representative of 5 independent experiments performed.

Transfectant	Percent Enhancement
2S	108$\pm$5
10S	119$\pm$10
30S	122$\pm$10
7.5R	194$\pm$24***
7.8R	204$\pm$40**
17.1R	168$\pm$11*
17.3R	186$\pm$5**

This demonstration that Nramp influences three independent pleiotropic effects of the gene previously associated with Ity/Lsh/Bcg function provides definitive evidence that Nramp is Ity/Lsh/Bcg.

these Nramp regulated pleiotropic effects rely on intracellular signalling mediated by the generation of mitochondrially-derived reactive oxygen intermediates (ROI).

Table 4 demonstrates that respiratory burst and L-arginine uptake are inhibited in the presence of the mitochondrial electron transport inhibitors rotenone (0-40 μM; inhibits complex I → ubiquinone) or thenoyltrifluoroacetone (TTFA; 0-400 μM; inhibits complex II → ubiquinone). Concentrations of inhibitors were based on previous studies (Ref. 61) examining the role of mitochondrially-derived ROI on apoptosis and the gene-inductive effects of TNFa in fibroblasts, and were not observed to have toxic effects on the RAW264.7-derived transfectant lines.

Table 4. L-arginine uptake experiments were performed in the presence of the radical scavengers nordihydroguaiaretic acid (0-40 μM) and butylated hydroxyanisole (0-400 μM). Respiratory burst and L-arginine uptake experiments were also carried out in the presence of the mitochondrial electron transport inhibitors rotenone (0-40 μM; inhibits complex I → ubiquinone) or thenoyltrifluoroacetone (TTFA; 0-400 μM; inhibits complex II → ubiquinone). Cells were allowed to adhere to microtitre wells for 1 hour prior to a 1 hour pretreatment with drugs before addition of activation agents for appropriate time periods. Results are presented for rotenone inhibition (percent of control) of L-arginine and RB for the resistant transfectant clone 7.5R examined after treatment with LPS/IFNγ.

Rotenone Concentration (μM)	L-arginine uptake	RB
0	100	100
5	66±16	52±3.7
10	79±17	33±4.4
20	73±7	16±5.4
40	74±11	11±3.7

These findings imply a role for Nramp in regulating mitochondrial function and the generation of reactive oxygen intermediates for signalling. Thus there are two ways in which Nramp may influence intracellular signalling for macrophage activation: (i) by influencing the generation of reactive oxygen intermediates from the mitochondrion; and (ii) by enhanced generation of nitric oxide. These studies of Nramp gene function bring together the decade of functional work demonstrating that Nramp regulates macrophage priming/activation for antimicrobial activity, with the many pleiotropic effects of the gene due to its role in regulating cell signalling events. The crucial significance of the putative SH3 binding domain in the function of the Nramp gene is that it regulates its function in response to priming/activation signals.

Nramp protein and antibody production.

On the basis of hydropathy plots the applicants have selected two structural domains that are not hidden by membrane and therefore are likely to be accessible within the intact/native protein conformation. Oligonucleotide primers to these two domains (N-terminal amino acids 1-82, C-terminal amino acids 514-548) were generated with restriction sites allowing the amplified products to be cloned in the appropriate reading frame in the pGEX series

enabling the induction of high level expression of fusion proteins that can be easily purified from bacterial lysates by affinity chromatography using glutathione agarose. Bound proteins can be released from the matrix under mild conditions such that the native conformation is maintained to improve antigenicity. This system has been employed to generate Nramp proteins of approximately 8.2 and 3.4 kd. from the N-terminal and C-terminal regions respectively, which have been used as antigens with the RIBI adjuvent to innoculate rabbits for production of polyclonal antibodies and rats for production of monoclonal antibodies. In order to ensure that antibodies raised will be specific to both murine and human Nramp/NRAMP proteins, these antibodies should be screened or affinity purified against peptides prepared on the basis of sequence information across these N-terminal and C-terminal regions used for production of the fusion protein. Specifically, against peptides DKSPPRLSRPSYGSISS; PQPAPCRETYLSEKIPIP; and GTFSLRKLWAFTGPGFLMSIAFLDPGNIESDLQ within the N-terminal region, and WTCCIAHGATFLTHSSHKHFLYGL in the C-terminal region.

Genomic Organization and Sequence
of Human NRAMP gene (Ref 62)

Genomic sequencing of NRAMP. A human yeast artificial chromosome (YAC) AM11/D3/14 (Ref 30), obtained from the ICRF library (available through the UK Human Gene Mapping Project HGMP Resource Centre, Huxton Hill, Cambridge CB10 1RQ, UK) by screening with a VIL1 probe (Ref 31) and containing the entire human NRAMP sequence (Ref 32), was sublconed into λEMBL3 (Stratagene Ltd, Cambridge, UK) and screened with the full-length murine *Nramp* cDNA λ8.1 (Ref 55). Two overlapping clones, λ3 and λB.1, containing the full-length NRAMP sequence, were digested with *Pst*I, sublconed into pBluescript II SK (Stratagene Ltd), and re-screened with the full-length murine cDNA probe (Ref 55). Exon positive clones were selected for sequence analysis, with

genomic sequence with [illegible]

Human cDNA sequence was obtained by reverse transcription (RT) and PCR amplification of RNA prepared from the human monocyte-derived THP1 cell line (Ref 33). Where appropriate, PCR products were cloned into the pCR vector (Invitrogen Corporation, Abingdon, UK) for sequence analysis from at least 2 independent clones. Clones corresponding to the 3' region were not originally isolated by screening with the murine cDNA. A fragment was generated by 3' rapid amplification of cDNA ends (RACE; (Ref 34)) from polydT adaptor primed THP1 cDNA. cDNA was amplified using the adaptor primer in combination with 2 nested primers selected from exon 13 (GTGCTGCCCATCCTCACG; GAGTTTGCCAATGGCCTG). A suitable genomic clone was prepared by amplification of a fragment from both λ3 and the YAC AM11/D3/14 using exon 13 primers and a primer (GGACGAGAAGGGAACTAG) designed from the 3' end of the RACE product. The 5' end of the RNA was mapped by 5' RACE involving RNA ligase-dependent ligation of a blocked anchor primer to the 3' end of random hexamer primed reverse transcribed THP1 RNA. Amplification using an anchor primer and two NRAMP-specific nested antisense primers (AAGAAGGTGTCCACAATGGTG, CGGTTTTGTGTCTGGGAT) yielded a single NRAMP product. The product was TA cloned and 3 clones subjected to sequence analysis to determine the transcriptional initiation site and sequence of the most proximal exon that failed to hybridise to any mouse cDNA probe. This facilitated further analysis of the 5' flanking region, the sequence for which was obtained from a 1.6 kb *PstI* fragment that contained sequence homologous to the 5' RACE product.

Analysis of sequence data. Nucleotide and amino acid sequence comparisons were made using the BESTFIT programme on-line to the CRC Resource Centre, UK. Amino acid sequences for murine and human NRAMP were aligned with yeast SMF1 and SMF2 (Ref 35) using the multiple sequence alignment program Clustal V (Ref 36).

exons 4 and 6 of human NRAMP, using RNA purified from peripheral blood mononuclear cells. This product spans the region of murine *Nramp* which carries the susceptibility mutation. PCR products were purified with a Qiagen PCR purification kit (Hybaid Ltd, Teddington, UK), and subjected to direct cycle sequence analysis using the CircumVent Thermal Cycle Dideoxy DNA Sequencing Kit (New England Biolabs, CP Laboratories, Bishop's Stortford, UK) with an internal sequencing primer (CATCTCTACTACCCCAAGGTGC). Direct cycle sequence analysis was performed on 19 individuals: 8 visceral leishmaniasis patients, 9 unaffected individuals taken from the same families, and 2 nonendemic British controls. Endemic samples were from Brazil (4 affecteds; 5 unaffecteds) and the Sudan (4 affecteds; 4 unaffecteds).

Primer design and PCR analysis of a 5' gt repeat using human genomic DNAs. PCR products of 780-794 bp were amplified from genomic DNA using primers located -365 bp 5' of the transcription start site (GAGGGGTCTTGGAACTCCA) and within intron 1 (CACCTTCTCCGGCAGCCC). This product was reamplified to generate 108-122 bp products using the 5' primer and an end-labelled (γ^{32}PdATP; ICN Biomedicals Ltd, Thame, UK) internal reverse primer TACCCCATGACCACACCC. The products were resolved by denaturing polyacrylamide gel electrophoresis and sized using a sequencing ladder. PCR products corresponding to different allelic forms were directly sequenced as described above.

Family linkage studies. A set of 36 multicase families of leprosy, tuberculosis and visceral leishmaniasis from our study site in Brazil (ref 37) were used to determine linkage between a polymorphic gt repeat in the 5' promoter region of human NRAMP and previously mapped 2q34-q35 markers (Refs 32, 37). Two-point linkage analyses were carried out between NRAMP and the markers (TNP1, IL8RB, VIL1, DES) using LINKAGE (Ref 38) on-line to the CRC Resource Centre. Gene frequencies for the NRAMP alleles were

Results

Referring to Figure 9, Clustal V multiple sequence alignment is shown for the deduced amino acid sequence for human NRAMP, murine *Nramp* clone λ8.1 (Ref 55), and the yeast mitochondrial proteins SMF1 and SMF2 (Ref 35). Residues showing 3/4 or 4/4 identities across the 4 proteins are shown in bold. For the NRAMP sequence: exon boundaries are indicated above the sequence; **PKC** indicates consensus sites (S/T-X-R/K) for protein kinase C phosphorylation; === indicates consensus sites for N-linked glycosylation; and putative membrane spanning domains ((Ref 2) are overlined and numbered on the sequence. * indicates cysteine residues conserved across all 4 proteins; . indicates conserved substitutions.

Referring to Figure 10(A), results of amino acid database searches for exon 2 are shown identifying a number of sequence matches with the Pro/Ser rich putative SH3 binding domain of NRAMP; + represents a conserved amino acid. Residues showing 4 or more identities are in bold. Multiple sequence alignments allowed for the generation of a consensus motif over this region as shown by double underlining. Also shown is the PKC site on S40, and tyrosine residues (*) on either side of the consensus motif. In Figure 10(B), Clustal V multiple sequence alignment is shown for human NRAMP, mouse *Nramp*, SMF1 and SMF2, and the expressed sequence tags (Ref 60) of *Oryza sativa* (rice; accession number d15268) and *Arabidopsis thaliana* (accession number z30530) genes, reading frames 1 and 2 respectively. Residues showing ≥4/6 identities across the 6 proteins are in bold. Membrane spanning domains 6 and 7 for NRAMP are overlined and numbered on the sequence. The 20 amino acid conserved transport motif (Ref 2) is indicated by double overlining. All 6 proteins show identities (similarities) of 7/20 (11/20) across the transport motif. * indicates cysteine residues conserved across all 6

promoter region human NRAMP sequence 5′ of the transcription start site. The transcription start site is located 148 bp 5′ of the ATG initiation codon, as indicated. Putative promoter region elements identified by inspection (indicated above the sequence) include: a possible Z-DNA forming dinucleotide repeat $t(gt)_5ac(gt)_5ac(gt)_9g$; 6 interferon-γ response elements; 3 W-elements; 1 AP1 site; 3 NFκB binding sites; and 9 purine-rich GGAA core motifs (2 on the antisense strand) for the myeloid-specific PU.1 transcription factor, two of which combine with imperfect AP1-like sites to create PEA-3 consensus motifs. Strings of heat shock transcription factor (HSTF) motifs (NGAAN or NTTCN) also occur across the 440 bp sequence (not marked).

Referring to Figure 12, two families are shown regregating for (a) alleles 2 and 3, or (b) alleles 1, 2 and 3 of the 5′ dinucleotide repeat polymorphism. Photographs below the families show autoradiographs of polymorphic PCR products (122 bp, 120 bp, and 118 bp for alleles 1 to 3, respectively) separated by denaturing polyacrylamide gel electrophoresis. Lanes from left to right on each photograph show individuals (a) I-2, II-1, II-2, II-3, II-4, II-5, II-6, III-1, III-2, III-3; and (b) I-1, I-2, II-1, II-2, III-1, III-2, III-3, III-4, III-5, III-6 as indicated on the pedigrees. Individual I-1 is not shown for family (a)

Sequence and genomic organization of human NRAMP. The sequencing of exon positive clones isolated by hybridization with a full-length cDNA allowed for the identification of the complete sequence (deposited with EMBL under accession numbers X82015 and X82016) of the human 2q homologue (NRAMP) of the murine chromosome 1 derived *Nramp* gene. Analysis of exon sequence from a region 440 bp 5′ of the transcriptional initiation site to the termination codon allowed for the complete exon-intron organization to be elucidated (Tables 5 and 6). Human NRAMP is encoded by 15 exons and, in constrast to the 548 amino acid

Kozak (Ref 16) consensus. The next, more distal codon found at M68 has a 2/6 Kozak consensus. However, we propose that like the murine macrophage form (Ref 55), the more proximal initiation codon will be utilised. This is reinforced by the striking (100%) sequence conservation for residues 51-67 (Fig. 9), indicating a requirement for the maintenance of sequence for function. The discrepancy in size between murine (548) and human (550) genes results from the inclusion of 3 additional residues within exon 2 causing a PTS duplication, with the non-duplicated form representing a rare variant in Brazilian (Ref 32) and British (unpublished data) pedigrees. In addition, the human gene exhibits a single amino acid deletion relative to the mouse within the poorly conserved last exon. Overall amino acid identity with murine *Nramp* was 86% (92% with conserved substitutions). Exons exhibiting highest sequence identity (100%) include exons 4, 6 and 7, with exon 11 displaying 98% identity. These exons encode TM1, the first extracellular domain, TM2 and TM3, and the conserved transport motif. It is of interest that TM2, containing the murine susceptibility-associated mutation (Refs 2, 56) is well conserved, suggesting that this domain plays an important functional role which cannot tolerate amino acid substitutions. NRAMP was aligned with murine *Nramp* and with the two yeast mitochondrial membrane proteins, SMF1 and SMF2, using the multiple sequence alignment program Clustal V (Fig. 9). SMF1 and SMF2, which show 49% identity (70% similarity) with each other, show 30% (57%) and 29% (53%) identities (similarities) with human NRAMP, respectively. This parallels the 30% (58%) and 30% (53%) identities (similarities) we reported (Ref 57) for murine *Nramp*. Regions of most striking sequence identity between all 4 proteins were found predominantly within the hydrophobic regions, although high identities were also found in exons 3, 4, 5 and 6, and for the conserved transport motif from exon 11. Within exon 6, the YAC-derived amino-acid human sequence exhibited a Gly at residue 172,

do not introduce negatively charged residues found in the susceptible allele of mice. As before (Refs 55, 32), matches with other proteins (Fig. 10) in the sequence databases were observed over exon 2 which contains a putative SH3 binding domain; and over the region of exon 11 containing the conserved binding-protein-dependent transport motif (Ref 2). The latter was highly conserved (7/20 identity; 11/20 similarity) in murine/human NRAMP, the yeast proteins, and in two expressed sequence tags from *Oryza sativa* (rice) and *Arabidopsis thaliana*. SMF1 and SMF2 do not demonstrate high identity over the proline/serine rich sequence of exon 2, but do have consensus (S/T-X-R/K) sequences (one in SMF1; two in SMF2) for PKC-dependent phosphorylation. Human NRAMP has two PKC consensus sites (in exons 2 and 3, Fig. 9) in this region, compared to three in the murine gene. The location of the distal site in SMF2 matches precisely with human NRAMP site 2/murine *Nramp* site 3, whereas the site in SMF1 is located 8 residues upstream. A pair of cysteine residues are conserved in all four genes: (i) in the first extracellular loop domain; and (2) in the third extracellular domain which also contains two sites for N-linked glycosylation in the human and murine genes. Charged residues are conserved across all 4 proteins within the transmembrane spanning domains 1,2,3,4, and 7 (Fig. 9), except for a Lys→Ser substitution in the first transmembrane domain of SMF1.

Analysis of the murine mutation site in visceral leishmaniasis patients and controls. To determine whether a mutation homologous to the murine disease susceptibility Gly→Asp mutation occurs in man, RT/PCR and direct cycle sequencing was performed on RNA from visceral leishmaniasis patients and controls from Brazil and the Sudan. All 19 human samples, whether from affected or unaffected individuals, encoded a Gly at this position.

site (Fig. 11). The transcription start site is located 148 bp 5′ of the ATG initiation codon. A series of predicted promoter region elements also occur 5′ of the transcription start site, including a possible Z-DNA forming (Refs 39, 40) dinucleotide repeat $t(gt)_5ac(gt)_5ac(gt)_9g$ located -317 to -274 bp 5′ of the transcription start site. On either side of the Z-DNA forming dinucleotide repeat are a series of matches to inducible promoter element consensus sequences. These include: 6 interferon-γ response elements, 1 x 3′→5′showing 8/8 matches to the consensus sequence $CT^G/_TG/_TANN^C/_T$ (Refs 41, 42), 3 x 5′→3′ showing 7/8 matches, 2 x 3′→5′ showing 7/8 matches; 3 W-elements (also known as H-, E-, W-, S-, or Z-boxes), 1 x 3′→5′ showing 8/8 matches to the consensus sequence $^A/_TGNA^C/_AC^C/_TG/_T$ (Ref 41), 2 x 5′→3′ with 7/8 matches; an AP1 site showing 6/7 matches to the consensus sequence TGACTCA (Ref 43); and 3 NFκB binding sites, 2 x 5′→3′ and 1 x 3′→5′, each showing 7/10 matches to the consensus sequence $GGG^G/_AC/A/_TT^C/_TC/_TCC$ (Ref 44). Nine purine-rich GGAA core motifs (2 on the antisense strand) for the myeloid-specific PU.1 transcription factor (Refs 45, 46) also occur across this region, two of which combine with imperfect AP1-like sites to create PEA3 motifs (Ref 47), and another two are juxtaposed. Strings of heat shock transcription factor (HSTF) motifs NGAAN or NTTCN; (Ref 48) were also present, although their order and phase are not consistent with currently functional elements. TATA, GC and CCAAT boxes were not found within the 440 bp 5′ flanking sequence.

Mapping of a polymorphic repeat in the 5′ promoter region. The presence of a gt repeat in the 5′ region of the YAC-derived NRAMP sequence stimulated further analysis of this region to determine whether a polymorphism was present in human population samples. Four alleles were observed in Brazilian families (Fig. 12): allele 1 = 122 bp; allele 2 = 120 bp; allele 3 = 118 bp; and allele 4 = 108 bp. Direct sequence analysis confirmed that the

allele 4 = $t(gt)_5ac(gt)_5ac(gt)_4g$. Gene frequencies determined on 72 genetically independent Brazilians were 0.021 (allele 1), 0.326 (allele 2), 0.646 (allele 3), and 0.007 (allele 4), providing an overall heterozygosity score of 0.476. Linkage analysis generated positive (>3) LOD scores (Table 7) for linkage between NRAMP and the four closest markers TNP1 (proximal) and IL8RB, VIL1, and DES (distal), consistent with physical mapping data (Ref 32) placing NRAMP 130 kb proximal to IL8RB, and confirming that this particular polymorphism occurs in the 2q35 copy of NRAMP rather than in a related sequence (Ref 49) mapping to a region in mice homologous to 6q27 in man.

Discussion

Genomic sequence analysis presented here demonstrates that the human NRAMP gene located on chromosome 2q35 has a genomic size of 12 kb and contains 15 exons. The amino acid sequence deduced from nucleotide sequencing of the 15 exons shows that, like murine *Nramp*, NRAMP encodes a polytopic integral membrane protein containing both a conserved transport motif (Ref 2) and a putative SH3 binding domain (Ref 55). Over the 20 amino acid transport motif, strong sequence identity (7/20 residues; 11/20 with conserved substitutions) was observed between NRAMP (*Nramp*), the two yeast proteins SMF1/2, and the expressed sequence tags from rice and *Arabidopsis*, suggesting that this is a functionally important motif among phylogenetically distinct organisms. Interestingly, these identities are higher than those reported (4/20 identity; 6/20 similarity) between murine *Nramp* and the nitrate transporter of *Aspergillus nidulans*, which led Vidal and coworkers (Ref 2) to hypothesise that *Nramp* might function in direct delivery of nitrates into the phagolysosomes of infected macrophages. The stronger identity observed here between the transport motif of NRAMP and the yeast mitochondrial proteins SMF1/2, together with the striking overall similarity between the yeast and human/murine genes, suggests that NRAMP may be a

possibly at the level of translocation (Ref 35). Complementation experiments with yeast mutants might therefore reveal more about the molecular mechanism of *Nramp* function. Sequence similarity between NRAMP(*Nramp*) and SMF1/2 was poor over the proline/serine rich putative SH3 binding domain. This is perhaps not unexpected as these are modular structures that occur in a variety of otherwise unrelated proteins involved in signalling and/or cytoskeletal attachment (Ref 55). Hence, this modular motif may be a recent addition to the NRAMP molecule related to its macrophage-restricted function, and we might expect that other more ubiquitously expressed NRAMP-like molecules will occur. A second *Nramp*-related sequence has already been mapped in the mouse (Ref 49), and others may be found.

Our major interest in analysing the human NRAMP gene was to provide the basis to screening multicase families for mycobacterial (tuberculosis and leprosy) and leishmanial infections. As a first step, we examined a small group of visceral leishmaniasis patients and their unaffected sibs to see whether a mutation similar to the murine susceptibility-associated mutation (Refs 2, 56) could be found. As might have been predicted, exon 6 encoding the second membrane spanning domain is highly conserved between murine and human sequences, as well as with the yeast genes, suggesting that this is a functionally important domain. No mutations were found within this region in the 19 human samples examined by direct cycle sequencing. Similarly, a polymorphic variant identified by us (Ref 32) in the putative SH3 binding domain occurred at very low frequency, suggesting that this too might be a region of the macrophage-expressed NRAMP molecule which, although recently acquired in evolutionary terms, may be critical to its function and intolerant to non-conservative substitutions.

The 440bp of promoter region sequence identified here is of

rather than cause structural changes to the molecule. Identification of PU.1 and PEA3/AP1-like response elements is consistent with haematopoietic-restricted gene expression (Refs 47, 50, 51). Although earlier studies (Refs 2, 55) suggest that murine *Nramp* is constitutively expressed in macrophages, the inducible promoter region elements identified in the human sequence suggest that expression may be regulated by macrophage priming/activation stimuli. In particular, interferon-γ and W-elements are common to other genes (e.g. MHC class II, (Ref 41); FcγRI (Ref 42); iNOS (Ref 52) inducible in macrophages. AP1 and NFκB sites also occur in the promoter regions of other macrophage-expressed proteins (e.g. tissue factor (Ref 43); iNOS (Ref 52) and are required for LPS and TNF inducibility, AP1 acting to stabilise and maintain NFκB activity (Ref 43). Given the many functional observations (reviewed in Refs 1, 57-59)) demonstrating that the *Ity/Lsh/Bcg* (candidate *Nramp*) phenotype is so closely allied to the interferon-γ/LPS macrophage activation pathway, it will be important to determine the functional relevance of these elements to tissue-specific expression of NRAMP in different macrophage subpopulations. This may be particularly relevant to previous observations demonstrating that the *Lsh* gene phenotype is differentially expressed in different macrophage subpopulations (Refs 53, 54), and that interaction with extracellular matrix elicits different levels of TNFα in bone marrow-derived macrophages from congenic resistant and susceptible mice (Ref 9). Although their order and phase were not consistent with currently functional elements, it was of interest that strings of HSTF elements were also found in the promoter region of human NRAMP. These may represent ancestral elements related to the mitochondrial activity/expression of the yeast SMF1 and SMF2 genes.

Another interesting feature of the 5′ flanking region of human NRAMP was the presence of a putative Z-DNA forming dinucleotide

regulatory signalling have been attributed to this form of DNA (reviewed Ref 39). It was particularly intriguing that a polymorphism in this repeat unit was observed in human genomic DNA samples. The fact that the putative Z-DNA forming repeat is flanked on either side by other promoter region response elements suggests that this polymorphism may be functionally important in determining gene expression, if not on the basis of its own role as a transcriptional regulator, at least because it will influence the juxtaposition of other response elements. The level of heterozygosity (0.476) in the Brazilian population studied here made this a useful marker for genetic linkage analysis between NRAMP and other 2q markers. However, the number of alleles was small compared to other repeat (e.g. microsatellite) polymorphisms, suggesting that the generation of further polymorphic variants across this repeat may not be tolerated in evolutionary terms. This polymorphism may therefore be of functional relevance in further analysis of the association between NRAMP and disease.

TABLES 5 and 6

Intron (4 flanking nucleotides)/exon(amino acids) boundaries and sizes (bp) for the 15 exons of human NRAMP identified by genomic sequence analysis of YAC-derived clones. Amino acid sequence identity with murine *Nramp* is shown for each exon.

TABLE 5

Exon Number	Size (bp)	Intron/exon boundaries				%AA Identity (Mouse)
EXON 1	155		Met Thr G ATG..145bp..ATG ACA G		gtga	50
EXON 2	143	acag	ly Asp Lys..(43aa)..Lys Pro GT GAC AAG.....AAA CCG		gtgg	68
EXON 3	123	acag	Gly Thr..(37aa)..Phe Lys GGC ACC .....TTC AAA		gtaa	95
EXON 4	120	acag	Leu Leu..(36aa)..Pro Lys CTT CTC.....CCT AAG		gtgg	100
EXON 5	107	tcag	Val Pro..(31aa)..Ala Gly Ar GTG CCC.....GCT GGA CG		gtac	91
EXON 6	71	tcag	g Ile Pro..(19aa)..Asn Tyr G A ATC CCA.....AAC TAC G		gtgg	100
EXON 7	68	gtag	ly Leu Arg..(18aa)..Tyr Gln GG CTG CGG.....TAT GAG		gtag	100
EXON 8	156	gcag	Tyr Val..(48aa)..Val Lys TAT GTG.....GTC AAG		gtag	88
EXON 9	159	gtag	Ser Arg..(49aa)..Ala Ala TCT CGA.....GCT GCG		gtga	87
EXON 10	90	gcag	Phe Asn..(26aa)..Gln Gly TTC AAC.....CAG GGG		gtga	80
EXON 11	120	gcag	Gly Val..(36aa)..Met Glu GGC GTG.....ATG GAG		gtag	98
EXON 12	150	ccag	Gly Phe..(46aa)..Leu Leu GGC TTC.....CTG CTG		gtag	94
EXON 13	74	ccag	Leu Pro..(20aa)..Asn Gly Le CTC CCG.....AAT GGC CT		gtga	84
EXON 14	154	ccag	u Leu Asn..(47aa)..Tyr Leu G CTG AAC.....TAC CTG		gtac	73
EXON 15	108	ccag	Val Trp..(34aa)..Ter GTC TGG.....TAG			67

Table 6

OVERLINED SEQUENCES REPRESENT REGIONS SELECTED FOR PRIMERS TO AMPLIFY INDIVIDUAL EXONS.

EXON 1

ATGTAAGAGGCAGGGCACTCGGCTGCGGATGGGTAACAGGGCGTGGGCTGGCACACTTACTT
GCACCAGTGCCCAGAGAGGGGGTGCAGGCTGAGGAGCTGCCCAGAGCACCGCTCACACTCCC
M T
AGAGTACCTGAAGTCGGCATTTCAATGACAGgtgagtagtggcccctagggacagagcctga
ttgggggggtggagtggaggagatcactaggctggtggagacttgaggaagcaagaaagccct
tggtcccctgtg

EXON 2 AMPLIFIED REGION

G D K G P Q R L S G S S
tcaccatgcttcatgggcccccacagGTGACAAGGGTCCCCAAAGGCTAAGCGGGTCCAGCT
Y G S I S S P T S P T S P G P Q Q A P P R
ATGGTTCCATCTCCAGCCCGACCAGCCCGACCAGCCCAGGGCCACAGCAAGCACCTCCCAGA
E T Y L S E K I P I P D T K P
GAGACCTACCTGAGTGAGAAGATCCCCATCCCAGACACAAAACCGgtgggacnctggaaact

ttctgggggct

EXON 3 AMPLIFIED REGION

aaggccagctgccaccatccctatacccacacccctcactctactcctcccacccccaacag
G T F S L R K L W A F T G P G F L M S I A
GGCACCTTCAGCCTGCGGAAGCTATGGGCCTTCACGGGGCCTGGCTTCCTCATGAGCATTGC
F L D P G N I E S D L Q L G P V A G F K
TTTCCTGGACCCAGGAAACATCGAGTCAGATCTTCAGCTNGGNCCNGTGGCGGGATTCAAAg

taactaagtcgggacctgagtgggacactt

EXON 4 AMPLIFIED REGION

L L W V L L W A
cctctctggctgaaggcctctccctgcctcctcacagCTTCTCTGGGTGCTGCTCTGGGCCA
T V L G L L C Q R L A A R L G V V T G K D
CCGTGTTGGGCTTGCTCTGCCAGCGACTGGCTGCACGTCTGGGCGTGGTGACAGGCAAGGAC
L G E V C H L Y Y P K
TTGGGCGAGGTCTGCCATCTCTACTACCCTAAGgtgagcttgggggcctggacagggagaa

ccactggccccaaaccccaaacagccattttcagcttccacga

[illegible]
V L W L T E A T M R F
CGTCCTCTGGCTGACCATCGAGCTAGCCATTGTGGGCTCCGACATGCAGGAAGTCATCGGCA

```
T  A  I  A  F  N  L  L  S  A  G  R
CGGCCATTGCATTCAATCTGCTCTCAGCTGGACGgtaccaccccagtgtccccaactcttca

ggcaggcagagaacagctgctgctacttccccccctaaccagtccctcccagagtctatttt

atcctgctgtcccctctgaagcagctgctgccctgttttccagaaatgtaaagtgacttgtc

taaagtcacacagatgtgagtcatgcaggaccccgggactgcag
```

EXON 6+7 AMPLIFIED REGION

```
                                                    I  P  L  W  G
agacccctggtcctggctgggctgacccgggccactctggtttcagAATCCCACTCTGGGGT
 G  V  L  I  T  I  V  D  T  F  F  F  L  F  L  D  N  Y
GGCGTCCTCATCACCATCGTGGACACCTTCTTCTTCCTCTTCCTCGATAACTACGgtgggtg
cacaccccacctcataggggagtggtggtggtgagggtgctgtactngggagaagggctctg
acatcgaacagcctgggagcgcacctgagctccctcactctcccctgggtgcctctagcgag
ttacttggacggctctcttcacctgtacatgggaaataatagcacagacttcagagggt---
----------------------32 bp----------------------------------
atagccatacgatgtgatgtcacagatttttcgtggnttggtttaggtttggtttggttct
        G  L  R  K  L  E  A  F  F  G  L  L  I  T  I  M  A  L
gctagtagGGCTGCGGAAGCTGGAAGCTTTTTTTGGACTCCTTATAACCATTATGGCCTTGA
T  F  G  Y  E
CCTTTGGCTATGAGgtaggaagccagtgctgcaacc
```

EXON 8

```
ggaagccagtgctgcaaccccactgtggacctcccaagatcattcctctcccttcctcctc
tggccgcgggnnnggggggggctggggtgggatggaggctgagaaatggtgaccgcggcg
                                                    Y  V  V  A  R
tggttgcgnggggcggggcttgtcctgaccaggctcctccctgcagTATGTGGTCGCCGTTC
P  E  Q  G  A  L  L  R  G  L  F  L  P  S  C  P  G  C  G  H  P
CTGAGCAGGGAGCGCTTCTTCGGGGCCTGTTCCTGCCCTCGTGCCCGGGCTGCGGCCACCCC
 E  L  L  Q  A  V  G  I  V  G  A  I  I  M  P  H  N  I  Y  L  H
GAGCTGCTGCAGGCGGTGGGACTTGTTGGCGCCATCATCATGCCCCACAACATCTACCTGCA
  S  A  L  V  K
CTCGGCCCTGGTCAAGgtgagcagaggggaggggaaagagacccccctcactcagtcggagcc
atgctggctccgcctccaannntggagccccct
```

EXON 9

```
ctgcagtgagccatgcattgcaccacggcactccagtctgggtgacagaacaaaacctgtct
ctaaaaaataaaataagtaagctggacacgtctgaggatggaacaaggtgagtgaaggagcg
tgtcaggacctgaggtagccagggacctcaaaggccagccttgcttcacccacacagtgctt
acagtggtaaggcctctgtggcaagaacagagatgtagaaaccatcggctgacctgaacctg
cccagactgccacgcagggcacttaagaaggtactgggctttggggagaacatagaagtgtg
agggtgggggacactgtggtggctctgagggactttggcacttccctctc

ACATGTACTTCCTGATTGAGGCCACCATCGCCCTGTCCGTCTCCTTTATCATCAACCTCTTT
 V  M  A  A  F  G  Q  A  F  Y  Q  K  T  K  Q  A  A
```

GTCATGGCTGCATTTGGGCAGGCCTTCTACCAGAAAACCAAGCAGGCTGCGgtgagacacac
tttcccccgcacctgaggccacacacgtactcatgtcctgtaagccttgccgaggaccctag
gcaatgcagctgagcccttctgagtctctgccctgatgatcttccctgttggcagatatcat
tcattcagcaaataatcattgagcatttgttatataccaagcacatcctagaccctggggat
acagcagtcaatgctacaaagacccagctctctgcag

EXON 10

tttggaaccctggtcagtgctaggcagtccagtttcccaaggctgaggntgctcctcac
F N I C A N S S L
tcacatcttccttctactgccctggtacccacagTTCAACATCTGTGCCAACAGCAGCCTCC
H D Y A K I F P M N N A T V A V D I Y Q G
ACGACTACGCTAAGATCTTCCCCATGAACAACGCCACCGTGGCCGTGGACATTTACCAGGGG

gtgagngngggtgggtggggagggcgtgacccagagaggcgcctcgggcagggccaccggtg
gtaccacactcgtccctgcag

EXON 11

ccgtggcactttaccgggggtgagcgcgggtgggtggggagggcgtgacccagagaggctc
G V I L G
cccgcctcgggcagggccaccggtcctaccacactcgtccctgcagGGCGTGATCCTGGGCT
C L F G P A A L Y I W A I G L L A A G Q S
GCCTGTTCGGCCCCGCGGCCCTCTACATCTGGGCCATAGGTCTCCTGGCGGCTGGGCAGAGC
S T M T G T Y A G Q F V M E
TCCACCATGACGGGCACCTACGCGGGACAGTTCGTGATGGAGgtagggcaggggggcgggcca
ggag

EXON 12

aaatgtttagtcttcagnaaccagctatgggatgggagttccccatttctccccacccatcc
cctcttgccacctagggacagagctgtcccagttcaacagtggaaaaacagagcatgccccc
agggataaatcggttgagggacatcagaggatctctcctctggaatccccagtcctgtctac
G F L R L R W S S F A
tcctcaccaaggagctcacccccaccccagGGCTTCCTGAGGCTGCGGTGGTCAAGCTTCGC
R V L L T R S C A I L P T V L V A V F R
CCGTGTCCTCCTCACCCGCTCCTGCGCCATCCTGCCCACCGTGCTCGTGGCTGTCTTCCGGG
D L R D L S G L N D L L N V L Q S L L
ACCTGAGGGACTTGTCGGGCCTCAATGATCTGCTCAACGTGCTGCAGAGCCTGCTGgtga

EXON 13

L P V A V L P I L T F T S M P T L M Q
ccagCTCCCGGTTGCCGTGCTGCCCATCCTCACGTTCACCAGCATGCCCACCCTCATGCAGG
E F A N G L
AGTTTGCCAATGGCCTgtgagtacccccttcccaagtgctggattgcatc

TCCATCATGGTGCTAGTCTGCACCATCAACCTCTACTTCGTGGTCAGCTATCTGCCCAGCCT

```
  P  H  P  A  Y  F  G  L  A  A  L  L  A  A  A  Y  L  G  L  S
GCCCCACCCTGCCTACTTCGGCCTTGCAGCCTTGCTGGCCGCAGCCTACCTGGGCCTCAGCA
T  Y  L
CCTACCTGgtacagtagggccaggggatgccttgggaatggatga
```

EXON 15 AMPLIFIED REGION

```
tgccttgggaatggatgattccccagaggtcttggcatctccccaattcatggttgcccctc
      V  W  T  C  C  L  A  H  G  A  T  F  L  A  H  S  S  H  H
ccccagGTCTGGACCTGTTGCCTTGCCCACGGAGCCACCTTTCTGGCCCACAGCTCCCACCA
   H  F  L  Y  G  L  L  E  E  D  H  K  G  E  T  S  G  *
CCACTTCCTGTATGGGCCTCCTTGAAGAGGACCACAAAGGGGAGACCTCTGGCTAGGCCCAC

ACCAGGGCTGGCTGGGGAGTGGCATGTATGACGT
```

TABLE 7.

Peak LOD scores for pairwise linkage analysis between NRAMP and previously mapped (Ref 37) 2q34 (TNP1) and 2q35 (IL8RB, VIL1, DES) markers calculated for 36 Brazilian families. RF = recombination fraction (M=F) at which the peak LOD score was obtained. N = number of families contributing to the analysis.

Marker intervals	N	Peak LOD Score	RF
TNP1-NRAMP	14	10.49	0.026
TNP1-IL8RB	9	6.02	0.032
TNP1-VIL1	15	9.84	0.001
TNP1-DES	19	11.45	0.046
NRAMP-IL8RB	11	3.56	0.072
NRAMP-VIL1	15	10.94	0.001
NRAMP-DES	20	8.94	0.051
IL8RB-VIL1	10	5.80	0.065
IL8RB-DES	12	10.03	0.035
VIL1-DES	14	9.47	0.059

REFERENCES

1. Blackwell, J. M. 1989. The macrophage resistance gene Lsh/Ity/Bcg. Res. Immunol. 140:767.

2. Vidal, S. M., D. Malo, K. Vogan, E. Skamene, and P. Gros. 1993. Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg. Cell 73:469.

3. Booker, G.W., I. Gout, A.K. Downing, P.C. Driscoll, J. Boyd, M.D. Waterfield, and I.D. Campbell. 1993. Solution structure and ligand-binding site of the SH3 domain of the p85α subunit of phosphatidylinositol 3-kinase. Cell 73:813.

4. Mayer, B.J. and D. Baltimore. 1993. Signalling through SH2 and SH3 domains. Trends Cell Biol. 3:8.

5. Gout, I., R. Dhand, I.D. Hiles, M.J. Fry, G. Panayotou, P. Das, O. Truong, N.F. Totty, J. Hsuan, G.W. Booker, I.D. Campbell and M.D. Waterfield (1993). The GTPase Dynamin Binds to and is Activated by a Subset of SH3 Domains. Cell. 75:25.

6. Morgenstern, J.P. and H. Land (1990) Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. Nucleic Acids Res. 18:3587.

7. Markowitz, D., S. Goft and A. Bank (1988) A safe packaging line for gene transfer: separating virol genes on two different plasmids. J. Virol. 62:1120.

8. Smith, D.B. and K.S. Johnson (1988) Single-step purification

9. Formica, S., T.I.A Roach and J.M. Blackwell (1994) Interaction with extracellular matrix proteins influences Lsh/Ity/Bcg (candidate Nramp) gene regulation of macrophage priming/activation for TNF-α and nitrite release. Immunology 82:42-50.

10. Otsu, M., I.D. Hiles, I. Goot, M.J. Fry, F. Ruiz-Larrea, G. Panayotou, A. Thompson, R. Dhand, J. Hsuan, N. Totty, A.D. Smith, S.J. Morgan, S.A. Courtneidge, P.J. Parker and M.D. Waterfield (1991). Characterization of two 85kd proteins that associate with receptor tyrosine kinsases, middle-T/pp60 (c-src) complexes, and PI3-kinase. Cell 65: 91.

11. Roach, T. I., A. F. Kiderlen, and J. M. Blackwell. 1991. Role of inorganic nitrogen oxides and tumor necrosis factor alpha in killing Leishmania donovani amastigotes in gamma interferon-lipopolysaccharide-activated macrophages from Lsh^s and Lsh^r congenic mouse strains. Infect. Immun. 59:3935.

12. Booker, G.W., A.I. Breeze, A.K. Downing, G. Panayotou, I. Gout, M.D. Waterfield and I.D. Campbell (1992) Structure of an SH2 domain of the p85 alpha subunit of phosphatidylinositol-3-OH kinase. Nature 358: 684.

13. Otto, E., M. Kunimoto, T. McLaughlin and V. Bennett. (1991) Isolation and characterization of cDNAs encoding human brain ankyrins reveal a family of alternatively spliced genes. J. Cell Biol. 114: 241.

14. Kyte, J. and R.F. Doolitle. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105.

15. Kozak, M. 1984. Point mutations close to the AUG initiator

16. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283.

17. Chen, M.S., R.A. Obar, C.C. Schroeder, T.W. Austin, C.A. Poodry, S.C. Wadsmworth and R.B. Vallee. 1991. Multiple forms of dynamin are encoded by shibire, a Drosophila gene involved in endocytosis. Nature 351:583.

18. Robinson, P. J., J. -M. Sontag, J. -P. Liu, E. M. Fykse, C. Slaughter, H. McMahon, and T. C. Südhof. 1993. Dynamin GTPase regulated by protein kinase C phosphorylation in nerve terminals. Nature 365:163.

19. Frielle, T., S. Collins, K.W. Daniel, M.G. Caron, R.J. Lefkowitz and B.K. Kobilka. 1987. Cloning of the cDNA for the human beta 1-adrenergic receptor. Proc. Natl. Acad. Sci. U.S.A. 84:7920.

20. Schaller, M.D., C.A. Borgman, B.S. Cobb, R.R. Vines, A.B. Reynolds and J.T. Parsons. 1992. $pp125^{FAK}$, a structurally distinctive protein-tyrosine kinase associated with focal adhesions. Proc. Natl. Acad. Sci. U.S.A. 89:5192.

21. Lipfert, L., B. Haimovich, M.D. Schaller, B.S. Cobb, J.T. Parsons and J.S. Brugge. 1992. Integrin-dependent phosphorylation and activation of the protein tyrosine kinase $pp125^{FAK}$. J. Cell Biol. 119:905.

22. Vallee, R.B. and H.S. Shpetner. 1993. Dynamin in synaptic dynamics. Nature 365:107.

23. Sadler, J., A.W. Crawford, J.W. Michelson and M.C. Beckerle. (1992) Zyxin and cCRP: Two interactive LIM domain proteins

Induction of early response genes KC and JE by mycobacterial lipoarabinomannans; regulation of KC expression in murine macrophages by *Lsh/Ity/Bcg* (candidate *Nramp*). Infect. Immun. 62: 1176-1184

25. Benit, L., M. Charon, L. Cocault, F. Wending and S. Gisselbrecht. 1993. The 'WS motif' common to v-mpl and members of the cytokine receptor superfamily is dispensable for myeloproliferative leukemia virus pathogenicity. Oncogene 8:787.

26. Kuramochi, S., T. Chiba, H. Amanuma, A. Tojo and K. Todokoro. 1991. Growth signal erythropoietin activates the same tyrosine kinases as interleukin 3, but activates only one tyrosine kinase as differentiation signal. Biochem. Biophysic. Res. Commun. 181:1103.

27. Ziegler, S.F., C.B. Wilson and R.M. Perlmutter. 1988. Augmented expression of a myeloid-specific protein tyrosine kinase gene (hck) after macrophage activation. J. Exp. Med. 168:1801.

28. English, B.K., J.N. Ihle, A. Myracle and T. Yi. 1993. Hck tyrosine kinase activity modulates tumor necrosis factor production by murine macrophages. J. Exp. Med. 178:1017.

29. Devereaux, J., P. Haeberli and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nuc. Acid Res. 12:387.

30. Larin Z, Monaco AP, Lehrach H (1991) Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc Natl Acad Sci USA 88: 4123-4127

31. Pringault E, Arpin M, Garcia A, Finidori J, Louvard D (1986)

32. White JK, Shaw M-A, Barton CH, et al (1994) Genetic and physical mapping of 2q35 in the region of NRAMP and IL8R genes: Identification of a polymorphic repeat in exon 2 of NRAMP. Genomics 24:295-302

33. Tsuchiya S, Yamabe M, Yamaguchi Y, Kobayashi Y, Konno T, Tada K (1980) Establishment and characterization of a human acute monocyte leukemia cell line (THP-1). Int J Cancer 26: 171-176

34. Edwards JBDM, Delort J, Mallet J (1991) Oligodeoxyribonucleotide ligation to single-stranded cDNAs: a new tool for cloning 5' ends of mRNAs and for constructing cDNA libraries by *in vitro* amplification. Nucleic Acids Res 19: 5227-5232

35. West AH, Clark DJ, Martin J, Neupert W, Hartl FU, Horwich AL (1992) Two related genes encoding extremely hydrophobic proteins suppress a lethal mutation in the yeast mitochondrial processing enhancing protein. J Biol Chem 267: 24625-24633

36. Higgins DG, Sharpe PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73: 237-244

37. Shaw M-A, Atkinson S, Dockrell H, et al (1993) An RFLP map for 2q33-q37 from multicase mycobacterial and leishmanial disease families: No evidence for an *Lsh/Ity/Bcg* gene homologue influencing susceptibility to leprosy. Ann Hum Genet 57: 251-271

38. Lathrop GM, Lalouel JM (1984) Easy calculations of lod scores and genetic risks on small computers. Am J Hum Genet 36: 460-465

39. Rich A, Nordheim A, Wang AH-J (1984) The chemistry and biology of left-handed Z-DNA. Ann Rev Biochem 53: 791-846

are highly dispersed in the human genome. Nature

41. Yang Z, Sugawara M, Ponath PD, et al (1990) Interferon gamma response region in the promoter of the human DPA gene. Proc Natl Acad Sci USA 87: 9226-9230

42. Pearse RN, Feinman R, Ravetch JV (1991) Characterization of the promoter of the human gene encoding the high-affinity IgG receptor: Transcriptional induction by gamma-interferon is mediated through common DNA response elements. Proc Natl Acad Sci USA 88: 11305-11309

43. Mackman N, Brand K, Edgington TS (1991) Lipopolysaccharide-mediated transcriptional activation of the human tissue factor gene in THP-1 monocytic cells requires both activator protein 1 and nuclear factor kB binding sites. J Exp Med 174: 1517-1526

44. Lenardo MJ, Baltimore D (1989) NF-kB: A pleiotropic mediator of inducible and tissue-specific gene control. Cell 58: 227-229

45. Klemsz MJ, McKercher SR, Celada A, Van Beveren C, Maki RA (1990) The macrophage and B cell-specific transcription factor PU.1 is related to the *ets* oncogene. Cell 61: 113-124

46. Karim FD, Urness LD, Thummel CS, et al (1990) The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. Genes and Development 4: 1451-1453

47. Cassady AI, Stacey KJ, Nimmo KA, et al (1991) Constitutive expression of the urokinase plasminogen activator gene in murine RAW264 macrophages involves distal and 5' non-coding sequences that are conserved between mouse and pig. Nucleic Acids Res 19: 6839-6847

48. Lis J, Wu C (1992) Heat shock factor. In: McKnight SL,

49. Dosik JK, Barton CH, Holiday DL, Krall MM, Blackwell JM, Mock BA (1994) An Nramp-related sequence maps to mouse chromosome 17. Mammalian Genome 5: 458-460

50. Feinman R, Qiu WQ, Pearse RN, et al (1994) PU.1 and an HLH family member contribute to the myeloid-specific transcription of the Fc-gamma-RIIIA promoter. EMBO J 13: 3852-3860

51. Zhang D-E, Hetherington CJ, Chen H-M, Tenen DG (1994) The macrophage transcription factor PU.1 directs tissue-specific expression of the macrophage colony-stimulating factor receptor. Mol Cell Biol 14: 373-381

52. Xie Q-w, Whisnant R, Nathan C (1993) Promoter of the mouse gene encoding calcium-independent nitric oxide synthase confers inducibility by interferon gamma and bacterial lipopolysaccaride. J Exp Med 177: 1779-1784

53. Crocker PR, Davies EV, Blackwell JM (1987) Variable expression of the murine natural resistance gene *Lsh* in different macrophage populations infected *in vitro* with *Leishmania donovani*. Parasite Immunol 9: 705-719

54. Davies EV, Singleton AM, Blackwell JM (1988) Differences in Lsh gene control over systemic Leishmania major and Leishmania donovani or Leishmania mexicana mexicana infections are caused by differential targeting to infiltrating and resident liver macrophage populations. Infect Immun 56: 1128-1134

55. Barton CH, White JK, Roach TIA, Blackwell JM (1994) NH_2-terminal sequence of macrophage-expressed natural resistance-associated macrophage protein (*Nramp*) encodes a proline/serine-rich putative *Src* homology 3-binding domain. J Exp Med 179: 1683-1687

sequence analysis of the mouse Nramp gene

to infection with intracellular parasites. Genomics 23: 51-61

57. Blackwell JM, Barton CH, White JK, et al (1994) Genetic regulation of leishmanial and mycobacterial infections: The *Lsh/Ity/Bcg* gene story continues. Immunol Lett 43: 99-107

58. Schurr E, Radzioch D, Malo D, Gros P, Skamene E (1991) Molecular genetics of inherited susceptibility to intracellular parasites. Behring Inst Mitt 88: 1-12

59. Schurr E, Malo D, Radzioch D, et al (1991) Genetic control of innate resistance to mycobacterial infections. Immunology Today 12: A42-A45

60. Boguski MS, Lowe TMJ, Tolstoshev CM (1993) dbEST - database for "expressed sequence tags". Nature Genet 4: 332-333

61. Schulze-Osthoff K. Beyaert R, Vandevoorde V, Haegeman G, Fiers W. (1993) Depletion of the mitochondrial electron transport abrogates the cytotoxic and gene-inductive effects of TNF. EMBO J. 12: 3095-3104.

62. Blackwell JM, Barton CH, White JK, Searle S, Baker A-M, Williams H, Shaw M-A (1994) Genomic Organisation and sequence of the human NRAMP gene: identification and mapping of a promoter region polymorphism. Mol. Med. (in press)

CLAIMS:

1. A natural resistance-associated macrophage protein having an N-terminal region comprising an SH3 binding domain.

2. A protein according to claim 1, wherein the N-terminal region further comprises one or more protein kinase C sites.

3. A protein according to claim 2, wherein the N-terminal region has two protein kinase C sites which flank the SH3 binding domain.

4. A protein according to any one of claims 1 to 3, wherein the SH3 binding domain comprises the SH3 binding motif PGPAPQPXPXR.

5. A protein according to claim 4, wherein the SH3 binding motif is PGPAPQPAPCR.

6. A protein according to claim 4 or claim 5, wherein the SH3 binding domain further comprises the polypeptide segment (S,A)PP(R,K)XSRPXXXS(I,V)XSX at the N-terminal end of the SH3 binding motif.

7. A protein according to claim 6, wherein the polypeptide segment is SPPRLSRPSYGSISSL.

8. A natural resistance-associated macrophage protein comprising the mouse amino acid sequence shown in Figure 9, optionally with mutations or deletions which do not substantially affect the activity thereof.

10. A protein according to claim 9, wherein the SH3 binding motif is PTSPTSPGPQQAPPRET.

11. A protein according to claim 9 or claim 10, wherein the SH3 binding domain further comprises the polypeptide segment GPQRLSGSSYGSISS at the N-terminal end of the SH3 motif.

12. A natural resistance-associated macrophage protein comprising the human amino acid sequence shown in Figure 9, optionally with mutations or deletions which do not substantially affect the activity thereof

13. A nucleotide sequence encoding a protein according to any one of the preceding claims.

14. A nucleotide sequence encoding a protein according to any one of claims 1 to 8, wherein the SH3 binding domain of the protein is encoded by the sequence comprising CCTGGCCCAGCACCTCAGCCAGCGCCTTGCCGG.

15. A nucleotide sequence according to claim 14, wherein the sequence encoding the SH3 binding domain further comprises the upstream region AGCCCCCCGAGGCTGAGCAGGCCCAGTTATGGCTCCATTT CCAGCCTG.

16. A nucleotide sequence encoding a protein according to any one of claims 1 to 3, or 9 to 12, wherein the SH3 binding domain of the protein is encoded by the sequence comprising CCG ACC AGC CCG ACC AGC CCA GGG CCA CAG CAA GCA CCT CCC AGA GAG ACC.

17. A nucleotide sequence according to claim 16, wherein the sequence encoding the SH3 binding domain further comprises the upstream region GGT CCC CAA AGG CTA AGC GGG TCC AGC TAT GGT TCC

17, which is a DNA sequence.

19. A nucleotide sequence according to claim 18, which is a cDNA sequence.

20. cDNA encoding natural resistance-associated macrophage protein, which comprises the nucleotide sequence shown in Figure 2, optionally with mutations or deletions which do not substantially affect the activity of the protein.

21. DNA encoding natural resistance-associated macrophage protein and comprising the following sequence, optionally with mutations or deletions which do not substantially affect the activity of the protein:

CCTGCGTCCTCATGATTAgtaatagcctccagggacctaatgggattcca
gtgggggtgacttgaggggcaaggaagatttagggtctctgtgggggct
cagctctgccagagcgacttcagtcaggcacttctgtggattaaacctgg
tggaggagacagagcatgggggtcaagcagctgagcgaggggcctcctgt
ctcacaaatctcctgactcaggggatttggattggagaagttctgttcct
cactgggagggaagtgattcttggaacctctgcttggcacataggtggac
ctgccagttgcggggagggaggtcgaggtcgtgggaggaggcaggtggct
tgaatcccaggctcttgaaagaagcacacacccacctagcatcctggggt
ccctgacagGTGACAAGAGCCCCCCGAGGCTGAGCAGGCCCAGTTATGGC
TCCATTTCCAGCCTGCCTGGCCCAGCACCTCAGCCAGCGCCTTGCCGGGA
GACCTACCTGAGTGAGAAGATCCCCATTCCCAGCGCAGACCAGgtaggga
tggtaggaatgtcctcagtgcttcccaggtcctaccggatccgagctcgg
accaag---300bp---
tcaagcttgagtgcatgtgtagctgtgtccattatagagcatgcgcgtgg
aggtcagaggacaacttgtgagagtcagctcacttctactgcgtgggttc
caactctggtggccttagcctctgagcccttctcggtccccattgccac
actctaagcagattcctaggctgtcgggccaaaccctgaaatagagttga
gtgactgagacctcagtggtccccagagagaagagcctgaagtatgagaa
gggtctggggagggaagagctgtagcagggaggtttaattacaacaaggt
ccccctcttgggactctgagaagcctgaaagaggcaggcaggtcatgtgc

ggcattgggtgggtgatggaggacagggttggtcagt

atacctgctgagttttagctgaggggatggtcaaggccagctgcatcca
tccaggagctaacatgacccgatctgcttgcacccccagGGTACATTCAG
CCTGAGGAAGCTGTGGGCGTTCACGGGGCCTGGTTTCCTCATGAGCATCG
CTTTCCTTGACCCGGGAAACATTGAGTCCGACCTTCAAGCTGGCGCTGTG
GCTGGGTTCAAAgtactgagtctgggccgccatgcttgctttgtggggag
cactttccttagctaggacaggggagaccccagttttccagagccggctg
catgggtggtttttctgaggataagctcctatcggggaggaaaaggaacc
ttggagaaacccctggagaaaggatgctgtagggtgttagtcttcccgcc
caatccccatcagagaggctgctctggctgagcatctcctctgtttcctc
acagCTCCTCTGGGTGCTGCTCTGGGCCACTGTGCTAGGTTTGCTGTGCC
AGCGGCTGGCTGCCCGGCT
GGGCGTGGTGACAGGCAAGGACTTGGGTGAAGTCTGCCATCTCTACTACC
CCAAGGTGCCCCGCATCCTCCTCTGGCTGACCATTGAGCTGGCCATTGTG
GGCTCAGATATGCAGGAAGTCATCGGGACGGCTATCTCCTTCAATCTGCT
CTCCGCTGGACGCATCCCGCTGTGGGGCGGTGTACTGATCACCATTGTGG
ACACCTTCTTCTTCCTCTTCTTGGATAACTATGGTTTGCGCAAGCTGGAA
GCTTTCTTCGGTCTCCTCATTACCATAATGGCTTTGACCTTCGGCTATGA
GTATGTGGTAGCACACCCTTCCCAGGGAGCGCTCCTTAAGGGCCTGGTGC
TGCCCACCTGTCCGGGCTGTGGGCAGCCCGAGCTGCTGCAGGCAGTGGGC
ATCGTCGGTGCCATCATCATGCCCCATAACATCTACCTGCACTCAGCCTT
GGTCAAGTCTAGAGAAGTAGACAGAACCCGCCGGGTGGATGTTCGAGAAG
CCAACATGTACTTCCTGATTGAGGCCACCATCGCCCTATCGGTGTCCTTC
ATCATCAACCTCTTTGTCATGGCTGTTTTTGGTCAGGCCTTCTACCAGCA
AACCAATGAGGAAGCGTTCAACATCTGTGCCAACAGCAGCCTCCAGAACT
ATGCTAAGATCTTCCCCAGGGACAATAACACTGTGTCAGTGGATATTTAT
CAAGGAGGTGTGATCCTAGGCTGTCTCTTTGGCCCTGCGGCCCTCTACAT
CTGGGCAGTAGGTCTCCTGGCAGCGGGGCAGAGTTCTACTATGACCGGCA
CCTATGCAGGACAGTTCGTGATGGAGGGTTTCCTTAAGCTGCGGTGGTCC
CGCTTCGCTCGGGTCCTTCTCACGCGCTCTTGCGCCATCCTGCCCACTGT
GTTGGTGGCTGTCTTCCGAGACCTGAAGGACCTGTCCGGCCTCAACGATC
TACTCAATGTTCTGCAGAGTCTACTGCTGCCCTTCGCTGTACTGCCCATT
TTGACTTTCACCAGCATGCCAGCTGTCATGCAGGAGTTTGCCAACGGCCG
GATGAGCAAAGCCATCACTTCGTGCATCATGGCGCTAGTCTGCGCCATCA
ACCTGTACTTTGTGATCAGCTACCTGCCCAGCCTCCCGCACCCTGCCTAC

[illegible]

GTGCAGGGTTCCGGGTGACCGCGGCATCCAGCAAGCAAAGAGGCAACAGG
GCAGACACAGCAGAGCAATTGGAGGTCCCCTACTGGCTTTCTGGATTACC
GGTTTCCAGTTTGGACAAGTGCTTTACCTCGGAATAATGACACCATTCTT
ATCACCACAACCTAAGAGACTTAAAAAACACAGTGCCTGGGGCGAGAGAT
GGCTCAGGTGTGAAGAACACTAGCCACCACCCTTTCAGAAGATGGGGATT
CAATTCCCAGCATCAACGTGGTGGCTTTCAACTGAAGGTGACTCCAGTTC
CCAGAACACCTCAAACAGAACTGCCACAACTCCATTGTCTCACTCCAGCT
CGTGGAAGATGAAGGGAGGAGTCCTAAAGAGTTCTAGGTCGGGTCTCTGG
AGAGACGGCTCAGCTGTTAAGAGCACCCGACTGCTCTTCCAGAGGTCCTG
AGTTCAATTCCCAGCAACCACATGGTGGCTCACAACCATCCATAATGGGA
TCCCTCTTCTGGTGTGTCTGAAGACAACAACAGTGTCCTCACATATATAA
AATAAATAAATCTTAAAAAAAAAAAAAAAAAAAAAAAAACTCGAG

22. A DNA sequence encoding a protein according to any one of claims 1 to 3 or 9 to 12, which sequence comprises one or more of the exons shown in Table 3, each of which is flanked by intron boundary regions.

23. A nucleotide sequence comprising the promoter region of the nucleotide sequence according to any one of claims 13 to 22, which promoter region includes a poly gt site.

24. A nucleotide sequence according to claim 23, wherein the poly gt site is of general formula $t(gt)_5ac(gt)_5ac(gt)_ng$, in which n=o or an integer.

25. A nucleotide sequence according to claim 13, which is an RNA.

26. Plasmid pBabeλ8.1 incorporating cDNA according to claim 20.

27. A retroviral vector construct incorporating a nucleotide sequence according to claim 19 or claim 20.

least a portion thereof.

29. A nucleotide primer pair according to claim 28, wherein the portion of the nucleotide sequence to be amplified is the poly gt site.

30. A nucleotide primer pair capable of hybridising to a portion of the nucleotide sequence of any one of claims 13 to 24, which nucleotide sequence encodes the N-terminal region of the protein which comprises or is upstream of the SH3 binding domain.

31. A nucleotide primer pair capable of hybridising to an exon as defined in claim 22, or the intron boundaries thereof, so as to permit amplification of at least a portion of the exon.

32. A nucleotide primer pair according to claim 31, wherein the exon is exon 2 of the human NRAMP gene.

33. A nucleotide primer pair according to claim 32, wherein the portion of the exon to be amplified comprises the sequence encoding the SH3 binding domain.

34. A nucleotide probe capable of hybridising to at least a portion of the nucleotide sequence of any one of claims 13 to 24, which nucleotide sequence encodes the N-terminal region of the protein which comprises or is upstream of the SH3 binding domain.

35. A nucleotide probe according to claim 34, which comprises a cDNA sequence.

36. A nucleotide probe capable of hybridising to the nucleotide sequence according to claim 23 or claim 24, or to at least a portion of the DNA sequence according to claim 22.

38. A polypeptide fragment of a protein according to any one of claims 1 to 12, which comprises at least a portion of the N-terminal region.

39. A polypeptide fragment of a protein according to any one of claims 1 to 12, which comprises an amino acid sequence selected from DKSPPRLSRPSYGSISS, PQPAPCRETYLSEKIPIP, GTFSLRKLWAFTGPGFLMSIAFLDPGNIESDLQ and WTCCIAHGATFLTHSSHKHFLYGL.

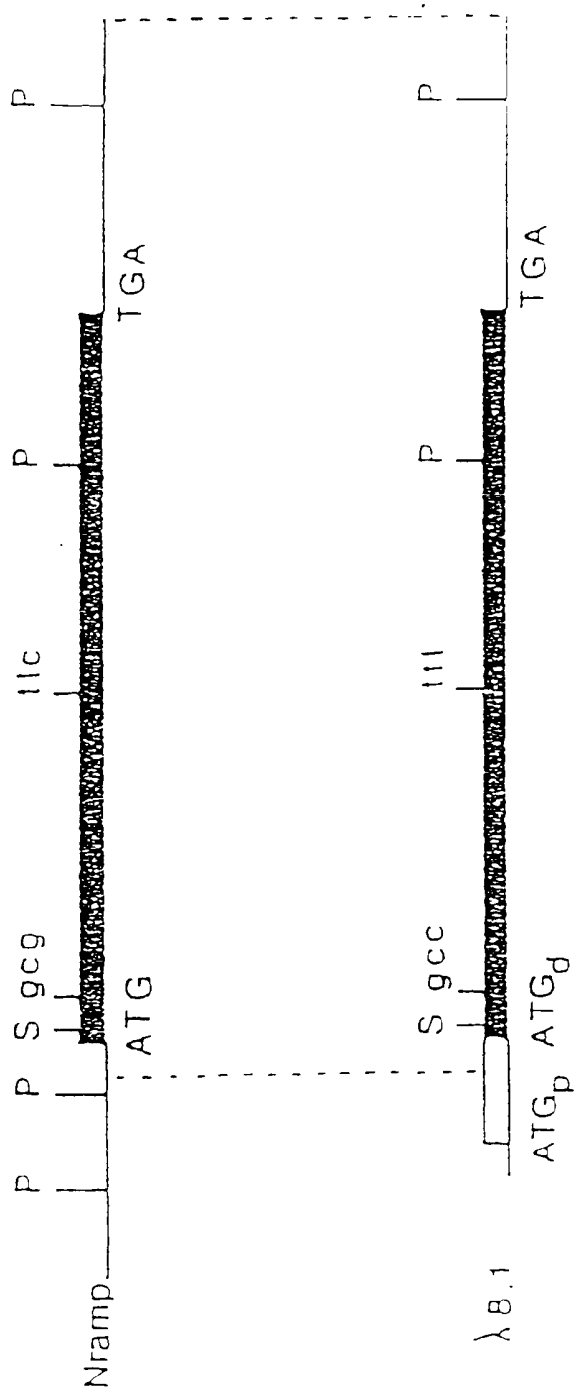
40. An antibody to a protein according to any one of claims 1 to 12.

41. An antibody to a polypeptide fragment according to claim 38 or claim 39.

42. An antibody according to claim 40 or claim 41, which is a monoclonal antibody.

43. Use of a primer pair according to any one of claims 28 to 33, in a diagnostic test to detect a polymorphism in the NRAMP gene.

44. Use of a probe according to any one of claims 34 to 37, in a diagnostic test to detect a polymorphism in the NRAMP gene.


P
P
TGA
P
P
TGA
IIc
III
P S gcg
ATG
S gcc
P
ATG$_p$ ATG$_d$
Nramp
λ 8.1

```
   1 GAATTCGGCACGAGGGAGCTAGTTGCCAGGCCTGGTGACCACACACAGAG
                          M  I  S  D  K  S  P  P  R  L  10
  51 TATCCTGCCGCCTGCGTCCTCATGATTAGTGACAAGAGCCCCCCGAGGCT
       S  R  P  S  Y  G  S  I  S  S  L  P  G  P  A  P  26
 101 GAGCAGGCCCAGTTATGGCTCCATTTCCAGCCTGCCTGGCCCAGCACCTC
       Q  P  A  P  C  R  E  T  Y  L  S  E  K  I  P  I  P  43
 151 AGCCAGCGCCTTGCCGGGAGACCTACCTGAGTGAGAAGATCCCCATTCCC
       S  A  D  Q  G  T  F  S  L  R  K  L  W  A  F  T  G  60
 201 AGCGCAGACCAGGGTACATTCAGCCTGAGGAAGCTGTGGGCGTTCACGGG
       P  G  F  L  M  S  I  A  F  L  D  P  G  N  I  E   76
 251 GCCTGGTTTCCTCATGAGCATCGCTTTCCTTGACCCGGGAAACATTGAGT
       S  D  L  Q  A  G  A  V  A  G  F  K  L  L  W  V  L  93
 301 CCGACCTTCAAGCTGGCGCTGTGGCTGGGTTCAAACTCCTCTGGGTGCTG
       L  W  A  T  V  L  G  L  L  C  Q  R  L  A  A  R  L 110
 351 CTCTGGGCCACTGTGCTAGGTTTGCTGTGCCAGCGGCTGGCTGCCCGGCT
       G  V  V  T  G  K  D  L  G  E  V  C  H  L  Y  Y  126
 401 GGGCGTGGTGACAGGCAAGGACTTGGGTGAAGTCTGCCATCTCTACTACC
       P  K  V  P  R  I  L  L  W  L  T  I  E  L  A  I  V 143
 451 CCAAGGTGCCCCGCATCCTCCTCTGGCTGACCATTGAGCTGGCCATTGTG
       G  S  D  M  Q  E  V  I  G  T  A  I  S  F  N  L  L 160
 501 GGCTCAGATATGCAGGAAGTCATCGGGACGGCTATCTCCTTCAATCTGCT
       S  A  G  R  I  P  L  W  G  G  V  L  I  T  I  V  176
 551 CTCCGCTGGACGCATCCCGCTGTGGGGCGGTGTACTGATCACCATTGTGG
       D  T  F  F  F  L  F  L  D  N  Y  G  L  R  K  L  E 193
 601 ACACCTTCTTCTTCCTCTTCTTGGATAACTATGGTTTGCGCAAGCTGGAA
       A  F  F  G  L  L  I  T  I  M  A  L  T  F  G  Y  E 210
 651 GCTTTCTTCGGTCTCCTCATTACCATAATGGCTTTGACCTTCGGCTATGA
       Y  V  V  A  H  P  S  Q  G  A  L  L  K  G  L  V  226
 701 GTATGTGGTAGCACACCCTTCCCAGGGAGCGCTCCTTAAGGGCCTGGTGC
       L  P  T  C  P  G  C  G  Q  P  E  L  L  Q  A  V  G 243
 751 TGCCCACCTGTCCGGGCTGTGGGCAGCCCGAGCTGCTGCAGGCAGTGGGC
       I  V  G  A  I  I  M  P  H  N  I  Y  L  H  S  A  L 260
 801 ATCGTCGGTGCCATCATCATGCCCCATAACATCTACCTGCACTCAGCCTT
       V  K  S  R  E  V  D  R  T  R  R  V  D  V  R  E  276
 851 GGTCAAGTCTAGAGAAGTAGACAGAACCCGCCGGGGTGGATGTTCGAGAAG
       A  N  M  Y  F  L  I  E  A  T  I  A  L  S  V  S  F 293
 901 CCAACATGTACTTCCTGATTGAGGCCACCATCGCCCTATCGGTGTCCTTC
       I  I  N  L  F  V  M  A  V  F  G  Q  A  F  Y  Q  Q 310
 951 ATCATCAACCTCTTTGTCATGGCTGTTTTTGGTCAGGCCTTCTACCAGCA
       T  N  E  E  A  F  N  I  C  A  N  S  S  L  Q  N  326
```

FIGURE 2 (1/2)

```
2101 CAAGGAGGTGTGATCCTAGGCTGTCTCTTTGGCCCTGCGGCCCTCTACAT
```

```
      W  A  V  G  L  L  A  A  G  Q  S  S  T  M  T  G   376
1151  CTGGGCAGTAGGTCTCCTGGCAGCGGGGCAGAGTTCTACTATGACCGGCA
      T  Y  A  G  Q  F  V  M  E  G  F  L  K  L  R  W  S   393
1201  CCTATGCAGGACAGTTCGTGATGGAGGGTTTCCTTAAGCTGCGGTGGTCC
      R  F  A  R  V  L  L  T  R  S  C  A  I  L  P  T  V   410
1251  CGCTTCGCTCGGGTCCTTCTCACGCGCTCTTGCGCCATCCTGCCCACTGT
      L  V  A  V  F  R  D  L  K  D  L  S  G  L  N  D     426
1301  GTTGGTGGCTGTCTTCCGAGACCTGAAGGACCTGTCCGGCCTCAACGATC
      L  L  N  V  L  Q  S  L  L  L  P  F  A  V  L  P  I   443
1351  TACTCAATGTTCTGCAGAGTCTACTGCTGCCCTTCGCTGTACTGCCCATT
      L  T  F  T  S  M  P  A  V  M  Q  E  F  A  N  G  R   460
1401  TTGACTTTCACCAGCATGCCAGCTGTCATGCAGGAGTTTGCCAACGGCCG
        M  S  K  A  I  T  S  C  I  M  A  L  V  C  A  I     476
1451  GATGAGCAAAGCCATCACTTCGTGCATCATGGCGCTAGTCTGCGCCATCA
      N  L  Y  F  V  I  S  Y  L  P  S  L  P  H  P  A  Y   493
1501  ACCTGTACTTTGTGATCAGCTACCTGCCCAGCCTCCCGCACCCTGCCTAC
      F  G  L  V  A  L  F  A  I  G  Y  L  G  L  T  A  Y   510
1551  TTTGGCCTTGTGGCTCTGTTCGCAATAGGTTACTTGGGCCTGACTGCTTA
      L  A  W  T  C  C  I  A  H  G  A  T  F  L  T  H     526
1601  TCTGGCCTGGACCTGTTGCATCGCCCACGGAGCCACCTTCCTGACCCACA
      S  S  H  K  H  F  L  Y  G  L  P  N  E  E  Q  G  G   543
1651  GCTCCCACAAGCACTTCTTATATGGGCTCCCTAACGAGGAGCAGGGAGGC
      V  Q  G  S  G  *                                    548
1701  GTGCAGGGTTCCGGGTGACCGCGGCATCCAGCAAGCAAAGAGGCAACAGG
1751  GCAGACACAGCAGAGCAATTGGAGGTCCCCTACTGGCTTTCTGGATTACC
1801  GGTTTCCAGTTTGGACAAGTGCTTTACCTCGGAATAATGACACCATTCTT
1851  ATCACCACAACCTAAGAGACTTAAAAAACACAGTGCCTGGGGCGAGAGAT
1901  GGCTCAGGTGTGAAGAACACTAGCCACCACCCTTTCAGAAGATGGGGATT
1951  CAATTCCCAGCATCAACGTGGTGGCTTTCAACTGAAGGTGACTCCAGTTC
2001  CCAGAACACCTCAAACAGAACTGCCACAACTCCATTGTCTCACTCCAGCT
2051  CGTGGAAGATGAAGGGAGGAGTCCTAAAGAGTTCTAGGTCGGGTCTCTGG
2101  AGAGACGGCTCAGCTGTTAAGAGCACCCGACTGCTCTTCCAGAGGTCCTG
2151  AGTTCAATTCCCAGCAACCACATGGTGGCTCACAACCATCCATAATGGGA
2201  TCCCTCTTCTGGTGTGTCTGAAGACAACAACAGTGTCCTCACATATATAA
2251  AATAAATAAATCTTAAAAAAAAAAAAAAAAAAAAAAAAAACTCGAG
```

FIGURE 2 CONT.
(2/2)

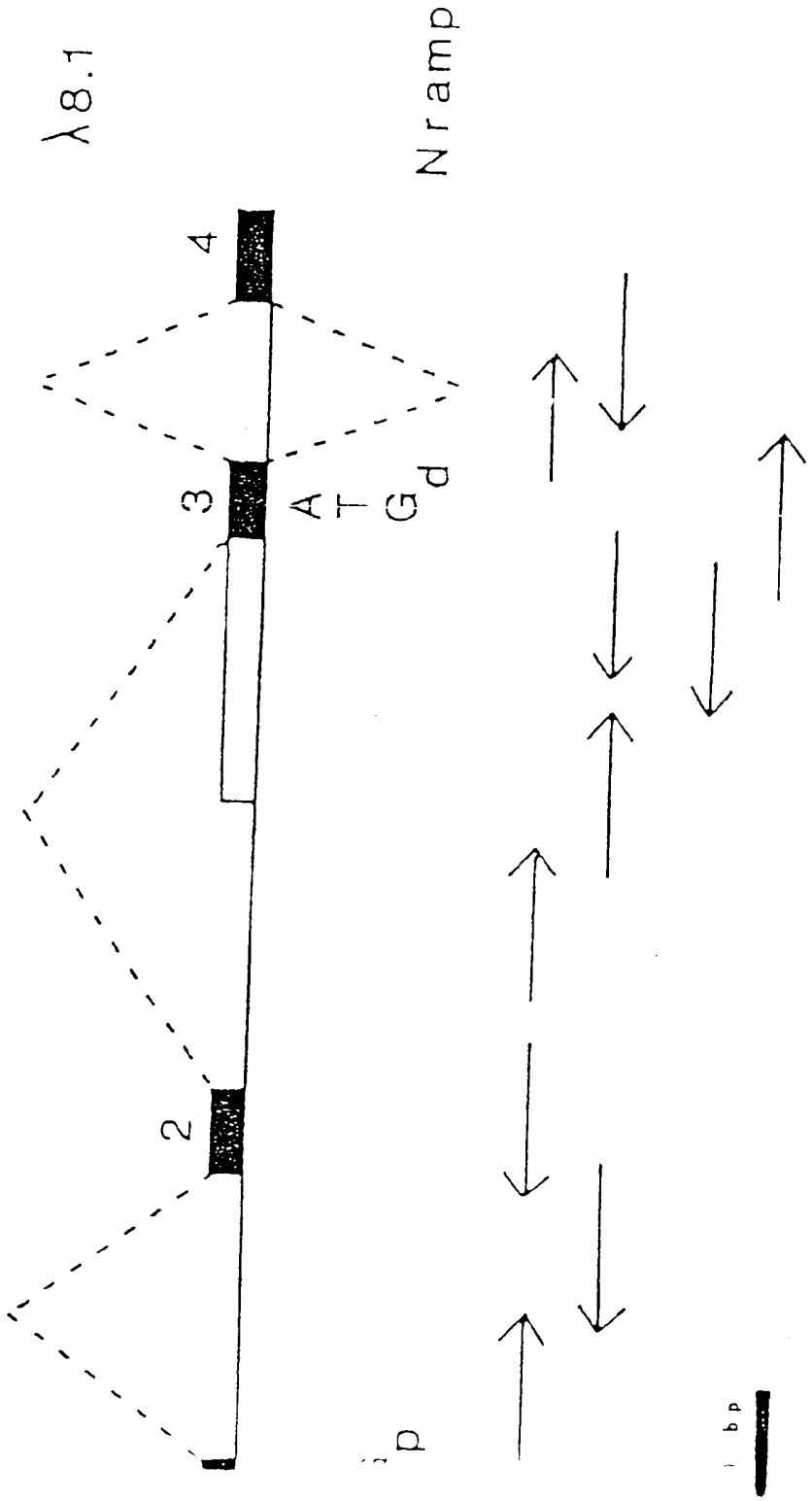


FIGURE 3

```
                  M  I                                                  2
   1  CCTGCGTCCTCATGATTAgtaatagcctccagggacctaatgggattcca

  51  gtgggggtgacttgaggggcaaggaagatttagggtctctgtgggggct
 101  cagctctgccagagcgacttcagtcaggcacttctgtggattaaacctgg
 151  tggaggagacagagcatgggggtcaagcagctgagcgagggcctcctgt
 201  ctcacaaatctcctgactcaggggatttggattggagaagttctgttcct
 251  cactgggagggaagtgattcttggaacctctgcttggcacataggtggac
 301  ctgccagttgcgggggagggaggtcgaggtcgtgggaggaggcaggtggct
 351  tgaatcccaggctcttgaaagaagcacacacccacctagcatcctggggt

                S  D  K  S  P  P  R  L  S  R  P  S  Y  G    16
 401  ccctgacagGTGACAAGAGCCCCCCGAGGCTGAGCAGGCCCAGTTATGGC

      S  I  S  S  L  P  G  P  A  P  Q  P  A  P  C  R  E    33
 451  TCCATTTCCAGCCTGCCTGGCCCAGCACCTCAGCCAGCGCCTTGCCGGGA

        T  Y  L  S  E  K  I  P  I  P  S  A  D  Q           47
 501  GACCTACCTGAGTGAGAAGATCCCCATTCCCAGCGCAGACCAGgtaggga
 551  tggtaggaatgtcctcagtgcttcccaggtcctaccggatccgagctcgg
 601  accaag---300bp---
 901  tcaagcttgagtgcatgtgtagctgtgtccattatagagcatgcgcgtgg
 951  aggtcagaggacaacttgtgagagtcagctcacttctactgcgtgggttc
1001  caactctggtggccttagcctctgagccccttctcggtccccattgccac

1051  actctaagcagattcctaggctgtcgggccaaaccctgaaatagagttga

1101  gtgactgagacctcagtggtccccagagagaagagcctgaagtatgagaa

1151  gggtctggggagggaagagctgtagcagggaggtttaattacaacaaggt

1201  ccccctcttgggactctgagaagcctgaaagaggcaggcaggtcatgtgc

1251  tggccagctgcagaggctgctgctgaacaggaccaacccagaaagcagag

1301  ccatagtgactcagcaaatggccctggtccctcgggggacgggcagcggt

1351  ggcattgggtgggtgatggaggacagggctggccagcctgactgaagaag

1401  atacctgctgagtttttagctgaggggatggtcaaggccagctgcatcca
                                                G  T  F  S  51
1451  tccaggagctaacatgacccgatctgcttgcaccccagGGTACATTCAG
        L  R  K  L  W  A  F  T  G  P  G  F  L  M  S  I    67
1501  CCTGAGGAAGCTGTGGGCGTTCACGGGGCCTGGTTTCCTCATGAGCATCG
      A  F  L  D  P  G  N  I  E  S  D  L  Q  A  G  A  V   84
1551  CTTTCCTTGACCCGGGAAACATTGAGTCCGACCTTCAAGCTGGCGCTGTG
       A  G  F  K                                          88
1601  GCTGGGTTCAAAgtactgagtctgggccgccatgcttgctttgtggggag
1651  cactttccttagctaggacaggggagaccccagttttccagagccggctg
1701  catgggtggtttttctgaggataagctcctatcggggaggaaaaggaacc
1751  ttggagaaaccccctggagaaaggatgctgtagggtgttagtcttcccgcc
1701  caatccccatcagagaggctgctctggctgagcatctcctctgtttcctc
          L  L  W  V  L  L  W  A  T  V  L  G  L  L  C    103
1851  acagCTCCTCTGGGTGCTGCTCTGGGCCACTGTGCTAGGTTTGCTGTGCC
      Q  R  L  A                                          107
```

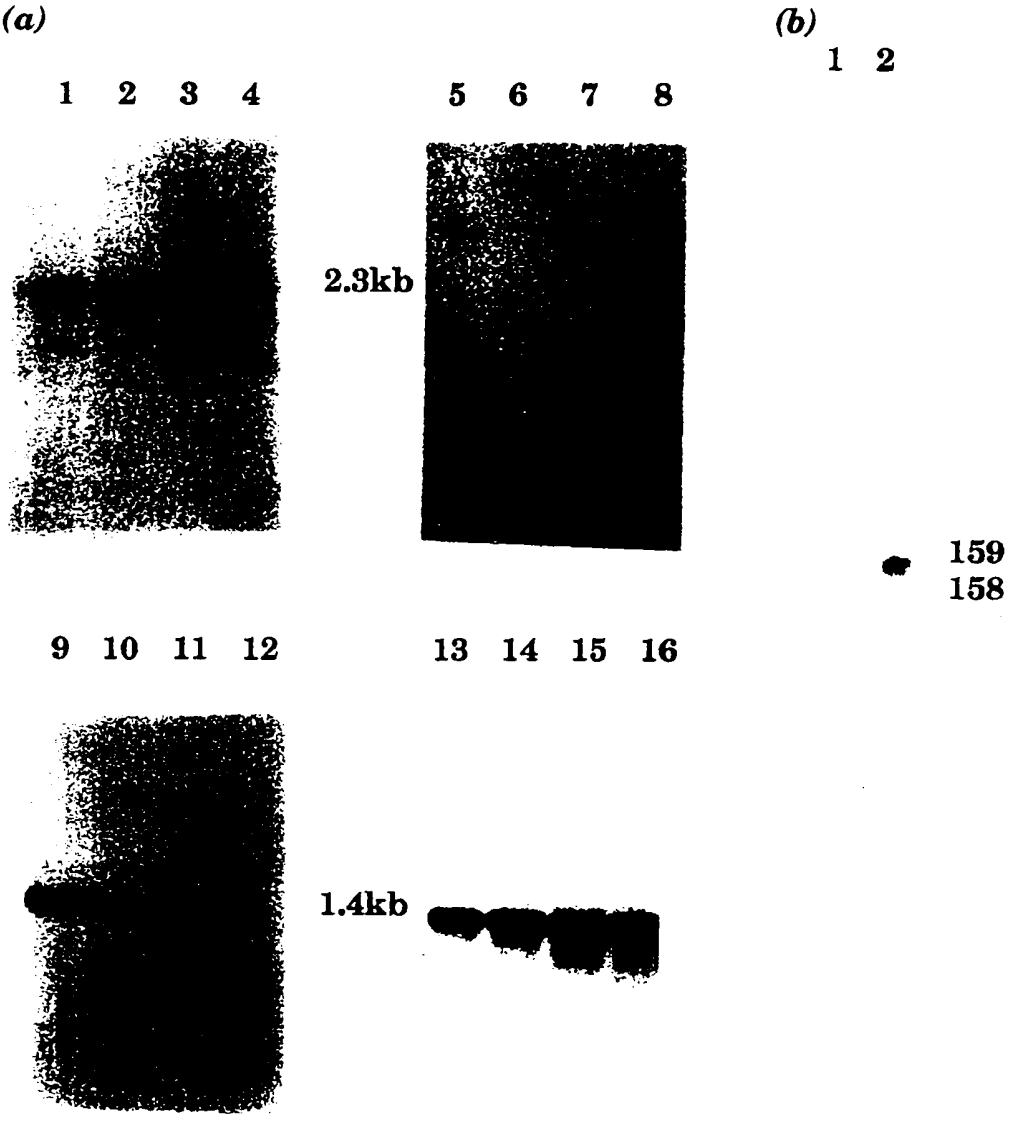
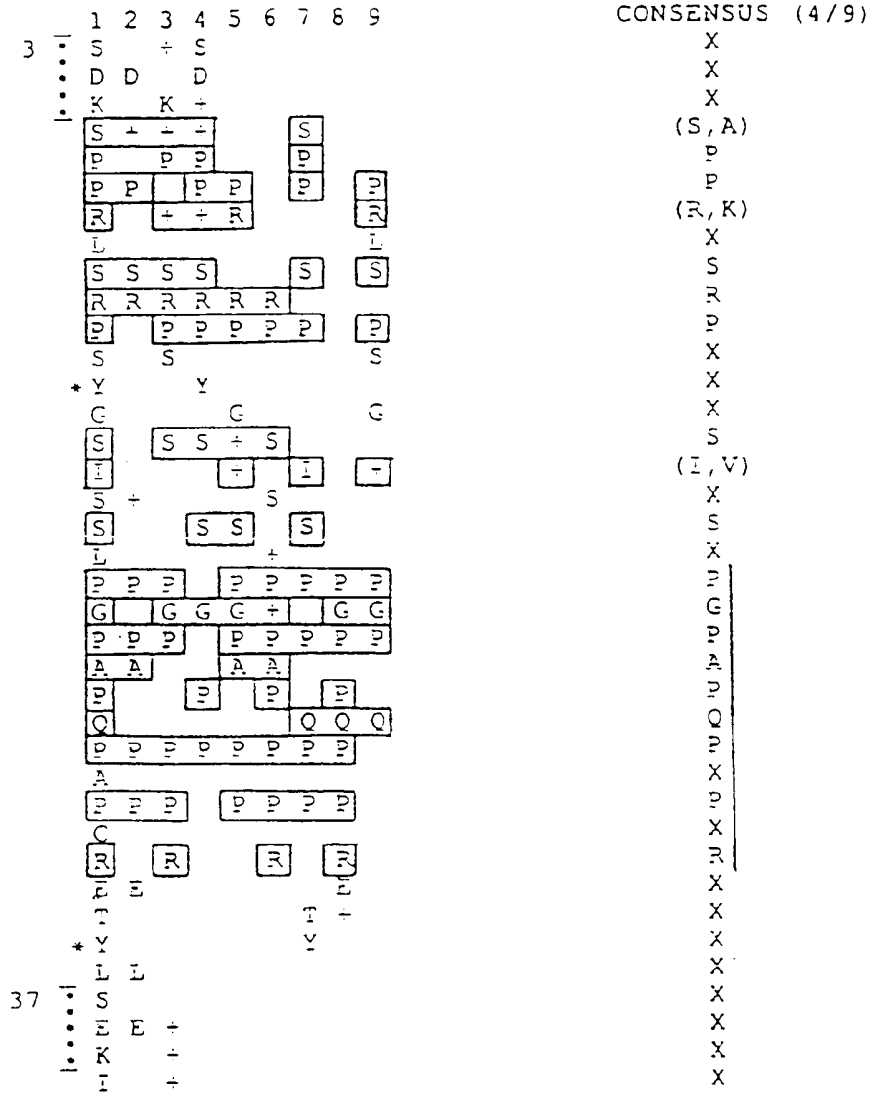


FIGURE 5



1.	NRAMP CLONE 8.1 DERIVED SEQUENCE	
2.	NIF-SPECIFIC REGULATORY PROTEIN	12/35 IDENTITIES (34%)
3.	GRANULINS PRECURSOR (AGROGRANIN)	13/38 IDENTITIES (34%)
4.	FAK PROTEIN TYROSINE KINASE	13/26 IDENTITIES (50%)
5.	TEGUMENT PROTEIN UL49	12/23 IDENTITIES (52%)
6.	BETA-1-ADRENERGIC RECEPTOR	12/21 IDENTITIES (57%)
7.	EXTENSIN (PRO RICH GLYCOPROTEIN)	14/30 IDENTITIES (46%)
8.	EBNA-1	9/13 IDENTITIES (69%)
9.	DYNAMIN (SHIBIRE PROTEIN)	11/20 IDENTITIES (55%)

Sequence alignment of exon 2 from human (TOP) and mouse Nramp.

```
       D   K   G   P   Q   R   L   S   G   S   S   Y   G   S
ag GT GAC AAG GGT CCC CAA AGG CTA AGC GGG TCC AGC TAT GGT TCC
   -- --- --- A-C --- -CG --- --G --- A-- C-- --T --- --C ---
   S   D   K   S   P   P   R   L   S   R   P   S   Y   G   S

   I   S   S   P   T   S   P   T   S   P   G   P   Q   Q   A
  ATC TCC AGC CCG ACC AGC CCG ACC AGC CCA GGG CCA CAG CAA GCA
  --T --- --- *** *** *** -T- C-T G-- --- -CA --T --- -C- --G
   I   S   S   *   *   *   L   P   G   P   A   P   Q   P   A

   P   P   R   E   T   Y   L   S   E   K   I   P   I   P   D
  CCT CCC AGA GAG ACC TAC CTG AGT GAG AAG ATC CCC ATC CCA GAC
  --- TG- C-G --- --- --- --C --- --- --- --- --- --T --C AG-
   P   C   R   E   T   Y   L   S   E   K   I   P   I   P   S

   T   K   P
  ACA AAA CCG gt
  G-- G-C -A-
   A   D   Q
```

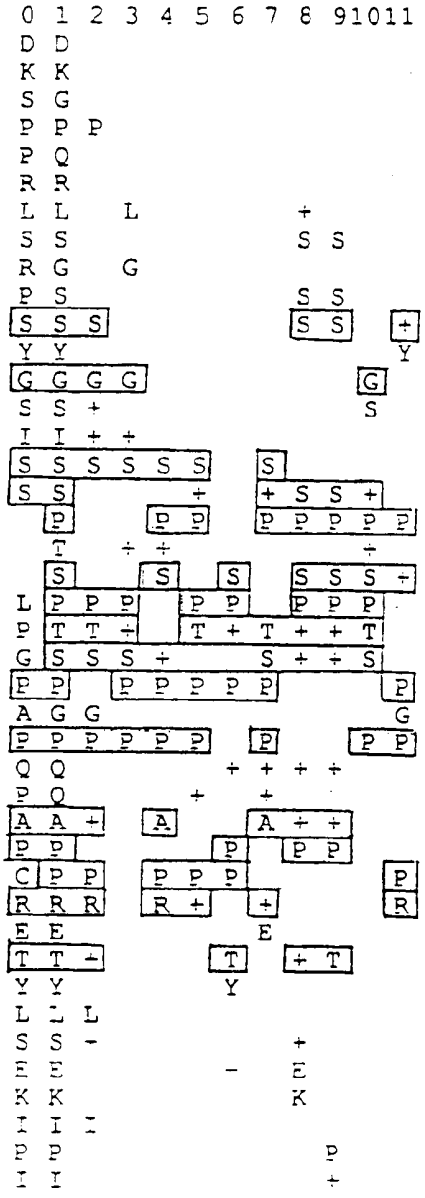
- indicates sequence identity

* indicates a gap introduce to maximise alignment

SEQUENCE IDENTITIES: nucleotide 115/147= 77%
amino acid 30/44= 68% (82% with conserved substitutions)

FIGURE 7

Alignment of human Nramp exon 2 with genbank amino acid database showing sequence identities and conserved substitutions (+).



		IDENTITY TO HUMAN NRAMP
0.	MOUSE NRAMP	72% (28/39)
1.	HUMAN NRAMP	
2.	MICROTUBULE-ASSOCIATED PROTEIN 4	32% (14/43)
3.	CYTOKINE RECEPTOR COMMON BETA CHAIN	45% (9/20)
4.	PHOSPHOLIPASE C-BETA 3	47% (8/17)
5.	ZYXIN REGION I (162 - 178)	41% (7/17)

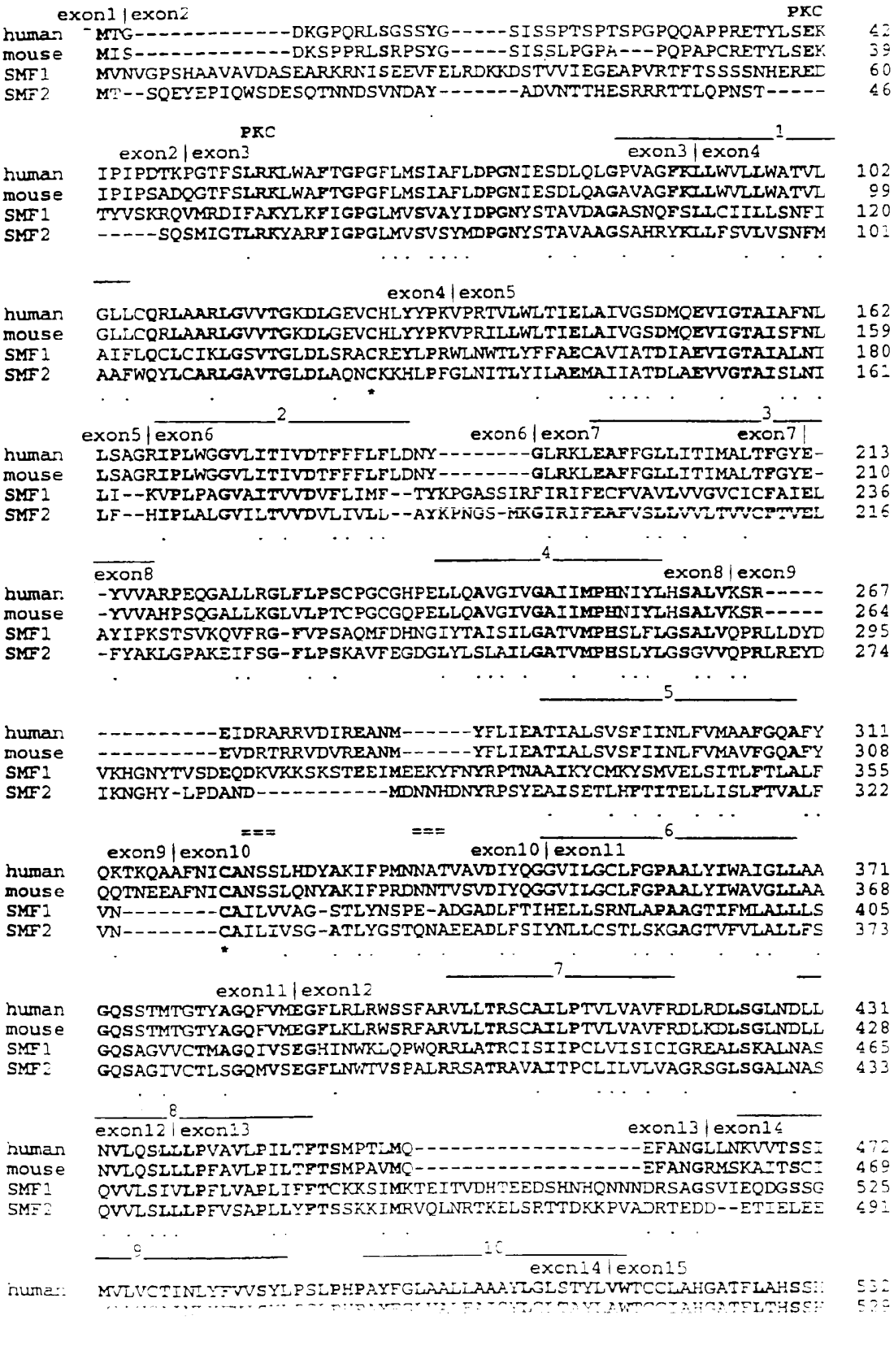
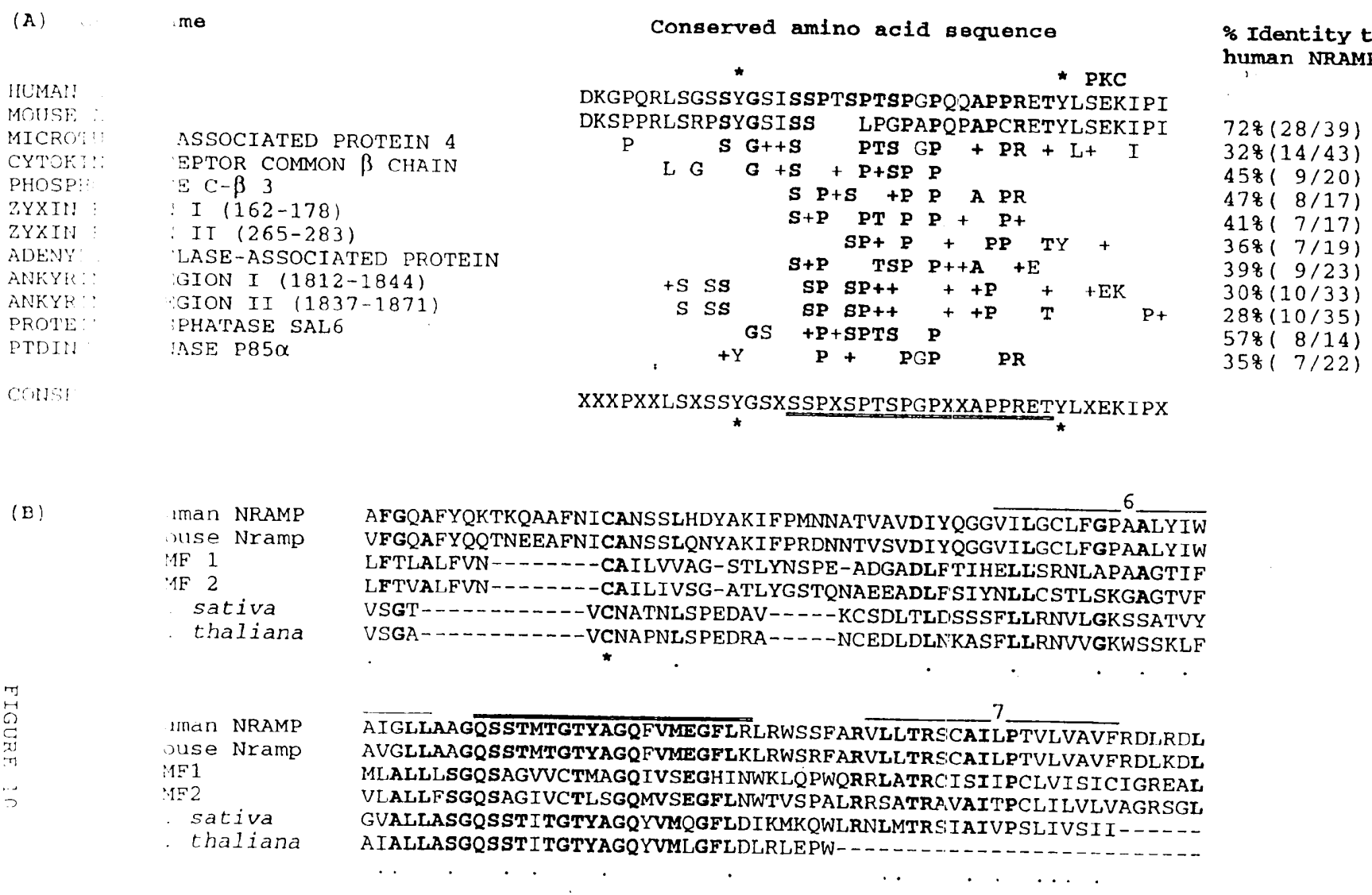

exon1|exon2
PKC
human MTG----------DKGPQRLSGSSYG-----SISSPTSPTSPGPQQAPPRETYLSEK 42
mouse MIS----------DKSPPRLSRPSYG-----SISSLPGPA---PQPAPCRETYLSEK 39
SMF1 MVNVGPSHAAVAVDASEARKRNISEEVFELRDKKDSTVVIEGEAPVRTFTSSSSNHERED 60
SMF2 MT--SQEYEPIQWSDESQTNNDSVNDAY-------ADVNTTHESRRRTTLQPNST----- 46
PKC
exon2|exon3
exon3|exon4
human IPIPDTKPGTFSLRKLWAFTGPGFLMSIAFLDPGNIESDLQLGPVAGFKLLWVLLWATVL 102
mouse IPIPSADQGTFSLRKLWAFTGPGFLMSIAFLDPGNIESDLQAGAVAGFKLLWVLLWATVL 99
SMF1 TYVSKRQVMRDIFAKYLKFIGPGLMVSVAYIDPGNYSTAVDAGASNQFSLLCIILLSNFI 120
SMF2 -----SQSMIGTLRKYARFIGPGLMVSVSYMDPGNYSTAVAAGSAHRYKLLFSVLVSNFM 101
exon4|exon5
human GLLCQRLAARLGVVTGKDLGEVCHLYYPKVPRTVLWLTIELAIVGSDMQEVIGTAIAFNL 162
mouse GLLCQRLAARLGVVTGKDLGEVCHLYYPKVPRILLWLTIELAIVGSDMQEVIGTAISFNL 159
SMF1 AIFLQCLCIKLGSVTGLDLSRACREYLPRWLNWTLYFFAECAVIATDIAEVIGTAIALNI 180
SMF2 AAFWQYLCARLGAVTGLDLAQNCKKHLPFGLNITLYILAEMAIIATDLAEVVGTAISLNI 161
exon5|exon6
exon6|exon7
exon7|
human LSAGRIPLWGGVLITIVDTFFFLFLDNY--------GLRKLEAFFGLLITIMALTFGYE- 213
mouse LSAGRIPLWGGVLITIVDTFFFLFLDNY--------GLRKLEAFFGLLITIMALTFGYE- 210
SMF1 LI--KVPLPAGVAITVVDVFLIMF--TYKPGASSIRFIRIFECFVAVLVVGVCICFAIEL 236
SMF2 LF--HIPLALGVILTVVDVLIVLL--AYKPNGS-MKGIRIFEAFVSLLVVLTVVCFTVEL 216
exon8
exon8|exon9
human -YVVARPEQGALLRGLFLPSCPGCGHPELLQAVGIVGAIIMPHNIYLHSALVKSR----- 267
mouse -YVVAHPSQGALLKGLVLPTCPGCGQPELLQAVGIVGAIIMPHNIYLHSALVKSR----- 264
SMF1 AYIPKSTSVKQVFRG-FVPSAQMFDHNGIYTAISILGATVMPHSLFLGSALVQPRLLDYD 295
SMF2 -FYAKLGPAKEIFSG-FLPSKAVFEGDGLYLSLAILGATVMPHSLYLGSGVVQPRLREYD 274
human ----------EIDRARRVDIREANM------YFLIEATIALSVSFIINLFVMAAFGQAFY 311
mouse ----------EVDRTRRVDVREANM------YFLIEATIALSVSFIINLFVMAVFGQAFY 308
SMF1 VKHGNYTVSDEQDKVKKSKSTEEIMEEKYFNYRPTNAAIKYCMKYSMVELSITLFTLALF 355
SMF2 IKNGHY-LPDAND----------MDNNHDNYRPSYEAISETLHPTITELLISLFTVALF 322
exon9|exon10
exon10|exon11
human QKTKQAAFNICANSSLHDYAKIFPMNNATVAVDIYQGGVILGCLFGPAALYIWAIGLLAA 371
mouse QQTNEEAFNICANSSLQNYAKIFPRDNNTVSVDIYQGGVILGCLFGPAALYIWAVGLLAA 368
SMF1 VN--------CAILVVAG-STLYNSPE-ADGADLFTIHELLSRNLAPAAGTIFMLALLLS 405
SMF2 VN--------CAILIVSG-ATLYGSTQNAEEADLFSIYNLLCSTLSKGAGTVFVLALLFS 373
exon11|exon12
human GQSSTMTGTYAGQFVMEGFLRLRWSSFARVLLTRSCAILPTVLVAVFRDLRDLSGLNDLL 431
mouse GQSSTMTGTYAGQFVMEGFLKLRWSRFARVLLTRSCAILPTVLVAVFRDLKDLSGLNDLL 428
SMF1 GQSAGVVCTMAGQIVSEGHINWKLQPWQRRLATRCISIIPCLVISICIGREALSKALNAS 465
SMF2 GQSAGIVCTLSGQMVSEGFLNWTVSPALRRSATRAVAITPCLILVLVAGRSGLSGALNAS 433
exon12|exon13
exon13|exon14
human NVLQSLLLPVAVLPILTFTSMPTLMQ----------EFANGLLNKVVTSSI 472
mouse NVLQSLLLPFAVLPILTFTSMPAVMQ----------EFANGRMSKAITSCI 469
SMF1 QVVLSIVLPFLVAPLIFFTCKKSIMKTEITVDHTEEDSHNHQNNNDRSAGSVIEQDGSSG 525
SMF2 QVVLSLLLPFVSAPLLYFTSSKKIMRVQLNRTKELSRTTDKKPVADRTEDD--ETIELEE 491
exon14|exon15
human MVLVCTINLYFVVSYLPSLPHPAYFGLAALLAAAYLGLSTYLVWTCCLAHGATFLAHSSH 532
human HHFLYGLLEEDHKG-ETS- 548
mouse KHFLYGLPNEEQGGVQGSG 548
SMF1 N--VYAIVQLG-MSHGDIS 575
SMF2 N--FYMLLGFT-TGKEVHL 549

FIGURE 9


(A)
me
Conserved amino acid sequence
% Identity to human NRAMP
HUMAN
MOUSE
MICROT
CYTOKI
PHOSP
ZYXIN
ZYXIN
ADENY
ANKYR
ANKYR
PROTE
PTDIN
ASSOCIATED PROTEIN 4
EPTOR COMMON β CHAIN
E C-β 3
I (162-178)
II (265-283)
LASE-ASSOCIATED PROTEIN
GION I (1812-1844)
GION II (1837-1871)
PHATASE SAL6
ASE P85α
CONSI
DKGPQRLSGSSYGSISSPTSPTSPGPQQAPPRETYLSEKIPI
DKSPPRLSRPSYGSISS LPGPAPQPAPCRETYLSEKIPI
72%(28/39)
32%(14/43)
45%(9/20)
47%(8/17)
41%(7/17)
36%(7/19)
39%(9/23)
30%(10/33)
28%(10/35)
57%(8/14)
35%(7/22)
XXXPXXLSXSSYGSXSSPXSPTSPGPXXAPPRETYLXEKIPX
(B)
uman NRAMP
ouse Nramp
MF 1
MF 2
sativa
thaliana
AFGQAFYQKTKQAAFNICANSSLHDYAKIFPMNNATVAVDIYQGGVILGCLFGPAALYIW
VFGQAFYQQTNEEAFNICANSSLQNYAKIFPRDNNTVSVDIYQGGVILGCLFGPAALYIW
LFTLALFVN--------CAILVVAG-STLYNSPE-ADGADLFTIHELLSRNLAPAAGTIF
LFTVALFVN--------CAILIVSG-ATLYGSTQNAEEADLFSIYNLLCSTLSKGAGTVF
VSGT----------VCNATNLSPEDAV-----KCSDLTLDSSSFLLRNVLGKSSATVY
VSGA----------VCNAPNLSPEDRA-----NCEDLDLNKASFLLRNVVGKWSSKLF
AIGLLAAGQSSTMTGTYAGQFVMEGFLRLRWSSFARVLLTRSCAILPTVLVAVFRDLRDL
AVGLLAAGQSSTMTGTYAGQFVMEGFLKLRWSRFARVLLTRSCAILPTVLVAVFRDLKDL
MLALLLSGQSAGVVCTMAGQIVSEGHINWKLQPWQRRLATRCISIIPCLVISICIGREAL
VLALLFSGQSAGIVCTLSGQMVSEGFLNWTVSPALRRSATRAVAITPCLILVLVAGRSGL
GVALLASGQSSTITGTYAGQYVMQGFLDIKMKQWLRNLMTRSIAIVPSLIVSII------
AIALLASGQSSTITGTYAGQYVMLGFLDLRLEPW----------
6
7
FIGURE 10

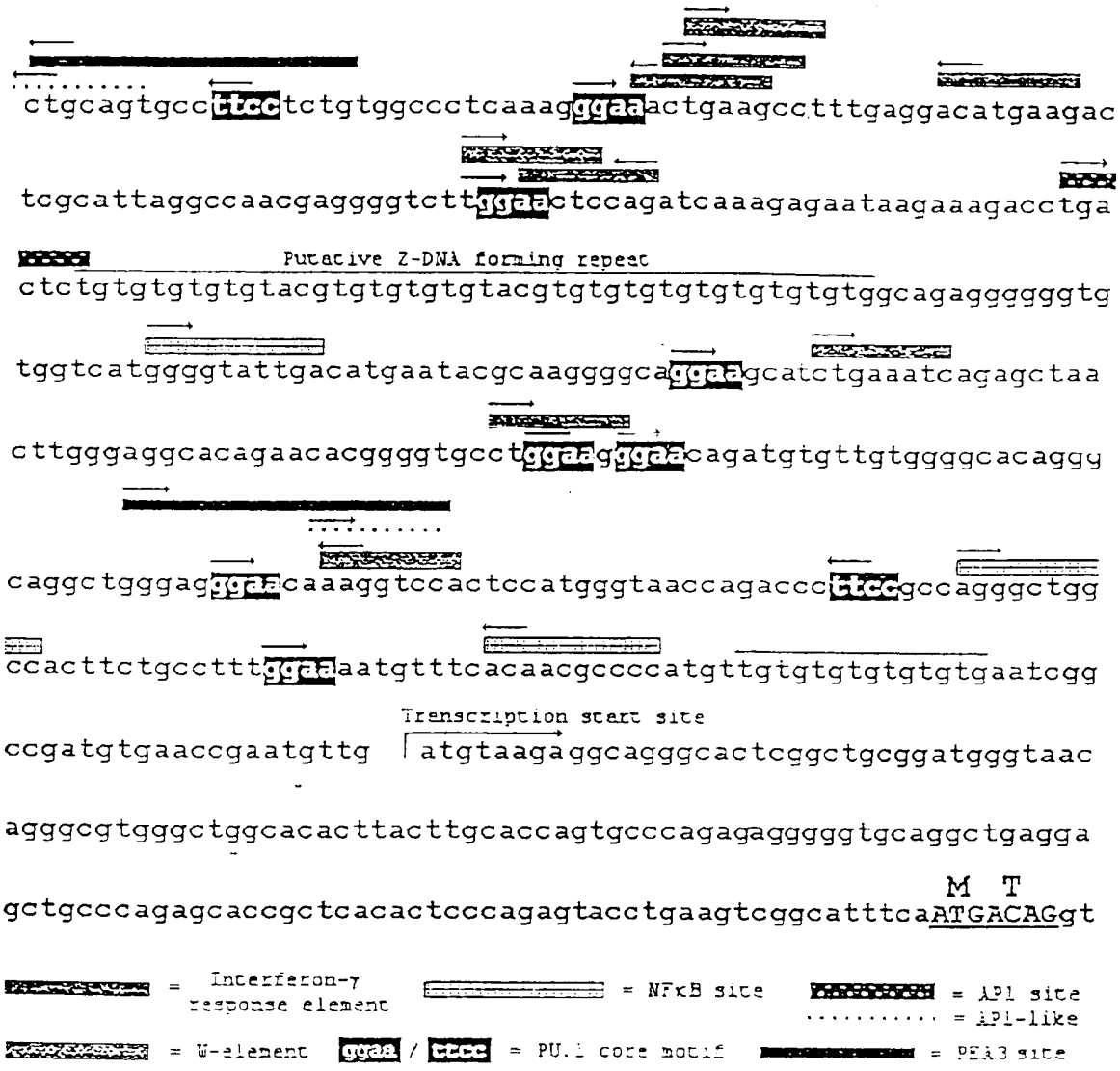
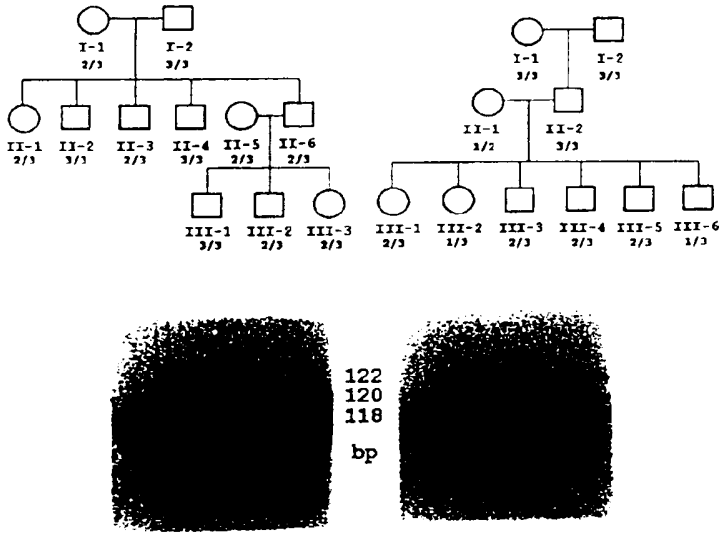

ctgcagtgcc ttcc tctgtggccctcaaag ggaa actgaagcctttgaggacatgaagac
tcgcattaggccaacgaggggtctt ggaa ctccagatcaaagagaataagaaagacctga
Putative Z-DNA forming repeat
ctctgtgtgtgtgtacgtgtgtgtgtacgtgtgtgtgtgtgtgtgtggcagaggggggtg
tggtcatggggtattgacatgaatacgcaaggggca ggaa gcatctgaaatcagagctaa
cttgggaggcacagaacacggggtgcct ggaa g ggaa cagatgtgttgtggggcacaggg
caggctgggag ggaa caaaggtccactccatgggtaaccagaccc ttcc gccagggctgg
ccacttctgcctttg ggaa aatgtttcacaacgccccatgttgtgtgtgtgtgtgaatcgg
Transcription start site
ccgatgtgaaccgaatgttg atgtaagaggcagggcactcggctgcggatgggtaac
agggcgtgggctggcacacttacttgcaccagtgcccagagaggggggtgcaggctgagga
M T
gctgcccagagcaccgctcacactcccagagtacctgaagtcggcatttca ATGACAG gt
= Interferon-γ response element
= NFκB site
= AP1 site
= AP1-like
= W-element
ggaa / ttcc = PU.1 core motif
= PEA3 site

FIGURE 11


I-1
2/3
I-2
3/3
II-1
2/3
II-2
3/3
II-3
2/3
II-4
3/3
II-5
2/3
II-6
2/3
III-1
3/3
III-2
2/3
III-3
2/3
I-1
3/3
I-2
3/3
II-1
3/2
II-2
3/3
III-1
2/3
III-2
1/3
III-3
2/3
III-4
2/3
III-5
2/3
III-6
1/3

122
120
118
bp

FIGURE 12

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 95/00095

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C12N15/86 C07K14/47 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CELL, vol. 73,no. 3, 7 May 1993 NA US, pages 469-485, SILVIA M. VIDAL ET AL. 'Natural resistance to infection with intracellular parasites: Isolation of a candidate for Bcg' see page 474, right column, paragraph 3 - page 477, left column, paragraph 1; figure 4 --- -/--	1-13

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report

European Patent Office
NL 2280 HV Rijswijk
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+ 31 70) 340-3016

Montero Lopez, B

Form PCT ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	THE JOURNAL OF EXPERIMENTAL MEDICINE, vol. 179,no. 5, 1 May 1994 pages 1683-1687, HOWARD BARTON ET AL. 'NH2-terminal sequence of macrophage-expressed natural resistance-associated macrophage protein (Nramp) encodes a proline/serine-rich putative Src homology 3-binding domain.' see summary see page 1684, left column, paragraph 4 - page 1687, right column, paragraph 1 ---	1-19,22
P,X	IMMUNOLOGY LETTERS, vol. 43,no. 1,2, 1994 pages 99-107, JENEFER M. BLACKWELL ET AL. 'Genetic regulation of leishmanial and mycobacterial infections: the Lsh/Ity/Bcg gene story continues' see page 101, right column, paragraph 1 - page 102, left column, paragraph 1; figure 5 -----	1-13,18, 19